THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with
publisher policies. Please scroll down to view the document
itself. Please refer to the repository record for this item and our
policy information available from the repository home page for
further information.

To see the final version of this paper please visit the publisher's website.
Access to the published version may require a subscription.

Author(s): Y Pokern, O Papaspiliopoulos, GO Roberts and AM Stuart
Article Title: Non parametric Bayesian drift estimation for one-dimensional diffusion processes
Year of publication: 2009
Link to published article:
http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2009/paper09-29

Publisher statement: None

Submitted to the Annals of Statistics

NONPARAMETRIC BAYESIAN DRIFT ESTIMATION FOR ONE-DIMENSIONAL DIFFUSION PROCESSES

By Yvo Pokern*, Omiros Papaspiliopoulos†, Gareth O. Roberts* and Andrew M. Stuart‡

We consider diffusions on the circle and establish a Bayesian estimator for the drift function based on observing the local time and using Gaussian priors. Given a standard Girsanov likelihood, we prove that the procedure is well-defined and that the posterior enjoys robustness against small deviations of the local time. A simple method for estimating the local time from high-frequency discrete time observations yielding control of the L^2 error is proposed. Complemented by a finite element implementation this enables error-control for a fixed random sample all the way from high-frequency discrete observation to the numerical computation of the posterior mean and covariance. An empirical Bayes procedure is suggested which allows automatic selection of the smoothness of the prior in a given family. Some numerical experiments extend our observations to subsets of the real line other than circles and exhibit more probabilistic convergence properties such as rates of posterior contraction.

1. Introduction.

1.1. *Setup.* This paper considers non-parametric estimation of drift functionals $b(\cdot)$ in one-dimensional Itô stochastic differential equations with constant diffusivity such as

$$dx = b(x)dt + dB, \quad x(0) = x_0. \tag{1}$$

Here, B refers to a standard Brownian motion and $b(\cdot)$ is the drift functional to be estimated, with anti-derivative denoted as $V' = b$. We shall consider the case where observations of x are available continuously on a finite time interval, and describe numerical methods to approximate to this ideal in the case where we have high-frequency observations.

Maximum likelihood estimation for this problem turns out to be ill-posed, and thus a Bayesian approach is natural, in which we can use priors which impose suitable smoothness on b so that the supports of prior and posterior distributions are sufficiently well-behaved. For technical reasons, we assume

*Department of Statistics, University of Warwick, Coventry, U.K.
†Department of Economics, Universitat Pompeu Fabra, Barcelona, Spain
‡Mathematics Institute, University of Warwick, Coventry, U.K.
AMS 2000 subject classifications: Primary 62G05; secondary 60K35

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

periodic boundary conditions on the prior for b, i.e. the SDE has the circle as a state space, which we parametrize by the interval $[0, 2\pi]$ with suitable identification of the end points. The compactness this provides offers most welcome simplifications of the proofs later. In the numerical part of the paper, we will exhibit how our methodology can be extended to bounded and unbounded intervals on the real line as state spaces.

1.2. *Likelihood.* To see why the likelihood we employ for our framework is reasonable, denote the measure on path space $\mathcal{C}([0,T],[0,2\pi])$ induced by Brownian motion on the circle as $\mathbb{Q}$. Similarly, denote the measure on the same space induced by (1) as $\mathbb{P}_b$. Then the Girsanov change of measure is given as

$$\frac{d\mathbb{P}_b}{d\mathbb{Q}} = \exp\left(-I[b]\right) \tag{2}$$

where the functional $I[b]$ is given as follows:

$$I[b] = \frac{1}{2}\int_0^T |b(x_t)|^2 dt - \int_0^T b(x_t) dx_t. \tag{3}$$

We now use Itô's formula for dV to eliminate the stochastic integral:

$$I[b] = \frac{1}{2}\int_0^T \Big(|b(x_t)|^2 + b'(x_t)\Big)\, dt - V(x_T) + V(x_0) - W\left(V(2\pi) - V(0)\right),$$

where we use W to denote the (signed) number of times the path wraps around the circle. Clearly, the winding number W is without influence whenever the potential V (and not just the drift function b) is periodic.

For statistical interpretation, it is natural to want to replace the time integral in (3) by a space integral in order to easily assess the *piecewise* influence of the likelihood on b. To this end, recall the local time of a stochastic process:

$$L_T(a) = \lim_{\epsilon \longrightarrow 0} \frac{1}{2\epsilon}\int_0^T \mathbf{1}_{(a-\epsilon, a+\epsilon)}(x_s) ds\ ,$$

If $L_T(\cdot)$ were differentiable, it would then be possible to rewrite (3) as

$$\begin{aligned} I[b] &= \frac{1}{2}\int_0^{2\pi} \Big(|b(a)|^2 + b'(a)\Big) L_T(a) da - V(x_T) + V(x_0) - W\left(V(2\pi) - V(0)\right) \\ &= \frac{1}{2}\int_0^{2\pi} \Big[|b(a)|^2 L_T(a) - b(a) L_T'(a) - 2b(a)\left(W + \tilde{\chi}(a; X_0, X_T)\right)\Big] da \end{aligned} \tag{4}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

We denote contributions due to initial and final condition using a modified indicator function

$$\tilde{\chi}(a; X_0, X_T) = \begin{cases} 1 & \text{if } X_0 < a < X_T \\ -1 & \text{if } X_T < a < X_0 \\ 0 & \text{otherwise} \end{cases} .$$

We aim to adopt the expression (4) as our log-likelihood. However unfortunately L_T is almost surely not differentiable at any point, so one of the complications of our approach will be to make clear mathematical sense out of this representation.

1.3. *Heuristic Bayesian Calculation.* To complete the Bayesian framework, a prior and resulting posterior measure are needed and it turns out that families of Gaussian processes

$$b \sim \mathcal{N}(b_0, \mathcal{A}^{-1})$$

with mean b_0 and precision operator $\mathcal{A}$ defined on a suitable function space H that takes periodicity into account are naturally conjugate within our context. In a first pass, we present heuristic calculations only which we will make rigorous in subsequent sections. We start by writing the Gaussian measure as a density with respect to (non-existing) Lebesgue measure on H:

$$p_0(b) \propto \exp\left(-\frac{1}{2}\int_0^{2\pi} (b-b_0)(a)(\mathcal{A}(b-b_0))(a)da\right) \tag{5}$$

Given an observation of a sample path x_t in continuous time, Bayes formula then yields the following formal density for the posterior measure:

$$\begin{aligned} \mathbb{P}\left(b|\{x_t\}_{t=0}^T\right) &\propto P(\{x_t\}_{t=0}^T|b)p_0(b) \\ &= \exp\Big(-\frac{1}{2}\int_0^{2\pi} Lb2 - bL' - 2b(W + \tilde{\chi}(\cdot; X_0, X_T))da \\ &\quad - \frac{1}{2}\int_0^{2\pi} (b-b_0)\mathcal{A}(b-b_0)da\Big) \end{aligned}$$

Completing the square in the exponent one finds that this is again a Gaussian with Mean

$$\widehat{b} = (\mathcal{A} + L_T)^{-1}\left(\frac{1}{2}L_T' + W + \tilde{\chi}(\cdot; X_0, X_T) + \mathcal{A}b_0\right) \tag{6}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

and covariance

$$Co = (\mathcal{A} + L_T)^{-1}, \tag{7}$$

where we abbreviate the informative part of the right hand side in (¡refeq:gaussMean) to f for future use:

$$f = \frac{1}{2}L_T' + W + \tilde{\chi}(\cdot; X_0, X_T)$$

Prior Elicitation. The conditional independence structure of the prior (and by inheritance the posterior) is specified infinitesimally through the operator $\mathcal{A}$, and this in turn determines the prior smoothness imposed on the function b. There is considerable latitude in choosing the $\mathcal{A}$, though we shall restrict ourselves to the consideration of priors with precision operators of the form

$$\mathcal{A} = \eta\Delta^k + \epsilon \tag{8}$$

for $\eta > 0$, $\epsilon \geq 0$ and $k \in \{1, 2\}$, where the Laplace operator Δ is simply the second derivative in this one-dimensional case: $\Delta = \frac{d^2}{da^2}$. The effect of the hyperparameters will be to enforce smoothness or low variability for large values of η, whereas large values of ϵ impose a stronger bias towards the constant zero function. The hyperparameter k affects the regularity of samples from prior and posterior measure which are in the Sobolev class, $H^{k-\frac{1}{2}-\varepsilon}$ for any $\varepsilon > 0$ (in any case, the SDE (1) still needs to make sense when using a typical sample from (5), see Subsection 5.0.2). So the choice $k = 1$ results in samples from the posterior which enjoy Brownian regularity, whereas $k = 2$ leads to the same regularity as that of integrated Brownian motion. This choice of precision operator not only results in familiar regularity but also in a Markovian structure so that it is possible to think of samples from the posterior as realizations of Brownian motion or its integral. Other, higher integer as well as fractional powers of Δ could be considered but are more cumbersome to implement for the discretisation methods considered.

The heuristic Bayesian calculations are made rigorous for the case $\eta = \epsilon = 1$ with $k = 2$ but this will carry over to any $\epsilon > 0$, $\eta > 0$ in a straightforward manner. We also perform numerical experiments in the cases of $k = 1$ and $\epsilon = 0$.

1.4. *Second Order, Gaussian Boundaries.*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

Introduction. In this subsection, we study a variation using, firstly, a second order precision operator instead of the fourth order operator studied in the analysis and, secondly, Robin boundary conditions are seen to arise from Gaussian (instead of periodic) boundary conditions. The Gaussian boundary conditions allow us to conceive of draws from the prior measure as realizations of the Brownian bridge with start and end points drawn independently from a Gaussian distribution. Draws from the posterior measure share the properties concerning the end points and could possibly be thought of as draws from a conditioned SDE.

While this subsection is mostly formal we believe that much of the analysis in the sequel could be made rigorous but renounce a complete exposition along the lines of the analysis presented in this paper for the sake of brevity.

Formal Calculation. In this paragraph, we construct a new prior measure implementing Gaussian boundary conditions. Consider multiplying another Gaussian density,

$$b_{\text{boundary}} \sim \exp\left(-\frac{1}{2\sigma^2}b^2(y) - \frac{1}{2\sigma^2}b^2(z)\right)$$

for the beginning and end points, y and z, with the improper prior

$$\tilde{\mu}_0 = \mathcal{N}(0, \Delta^{-1}) \oplus \lambda$$

where the Gaussian measure lives on $\dot{L}^2(y,z)$, the space of square integrable functions on (y,z) with average zero and we use Lebesgue measure on the space $\{a\mathbf{1} : a \in \mathbb{R}\}$ of constant functions.

We assume that the points y and z have been chosen such that the local time L_T is informative on $[y,z]$ but nearly zero outside. It then turns out that this translates into a Robin boundary condition for the PDE for the posterior mean, which we demonstrate in the following formal calculation for the posterior probability $\mu(b)$:

$$I[b] = -\log(\mathbb{P}(b|L_T, \text{prior})) \tag{9}$$

$$= \frac{1}{2}\int_y^z \eta(b')^2(a) + L_T(a)b^2(a) - b(a)L_T'(a) - 2b(a)\tilde{\chi}(a; X_0, X_T)da \tag{10}$$

$$+ \frac{1}{2\sigma^2}b^2(y) + \frac{1}{2\sigma^2}b^2(z) + \text{const} \tag{11}$$

To find a critical point of this functional, take its variation for small deviations εv from b and perform a partial integration:

$$I[b+\varepsilon v] - I[b] = \varepsilon\left[\int_y^z \eta b'v' + L_T bv - \frac{1}{2}vL_T - v\tilde{\chi}da + b(y)v(y) + b(z)v(z)\right] + \mathcal{O}(\varepsilon^2).$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

Now perform a partial integration in the first integral to find that the new PDE to be solved reads

$$-\eta b'' + L_T b = \frac{1}{2} L_T' + \tilde{\chi}, \quad b(y) = \sigma^2 \eta b'(y),\ b(z) = \sigma^2 \eta b'(z) \tag{12}$$

The new boundary condition relating the value of the function at the boundary to the value of its derivative is known as a Robin boundary condition.

1.5. *Estimating Smoothness Hyperparameters.* It is possible to employ an empirical Bayes framework, where the constants η and ϵ in the prior (8) are treated as hyperparameters taking values in $(0,\infty) \times [0,\infty)$. As the numerical experiments will highlight, performance of the estimator depends on the value of these constants, and while it is possible to extend to a hierarchical prior, we will simply employ maximization of the integration constant $P(\{x_t\}_{t=0}^T)$, i.e. we will require

$$\eta = \arg \max_{\eta \in (0,\infty), \epsilon \in [0,\infty)} \int_{L^2} P(\{x_t\}_{t=0}^T | b) p_0(db)$$

leaving the order parameter k fixed. The integral can be computed formally again by completing the square and turns out to be

$$\int_{L^2} P(\{x_t\}_{t=0}^T | b) p_0(db) = \sqrt{|\mathrm{Id} + \mathcal{A}^{-1}(\eta,\epsilon) L_T|^{-1}} \cdot \exp\left(-\frac{1}{2} \int_0^{2\pi} \left[-\left(\mathcal{A}(\eta,\epsilon) b_0 + f\right) \left(\mathcal{A}(\eta,\epsilon) + L_T\right)^{-1} \left(\mathcal{A}(\eta,\epsilon) b_0 + f\right) + b_0 \mathcal{A}(\eta,\epsilon) b_0 \right] da \right). \tag{13}$$

Although this computation is entirely formal, it is possible to evaluate the resulting expression numerically and investigate robustness of the results under refinement of resolution. It should be noted that there are no guarantees that a maximizer will exist in the specified region $(\eta,\epsilon) \in (0,\infty) \times [0,\infty)$, and we observe that, in fact, for some datasets, the maximum is attained at the boundary, while for other datasets reasonable and robust values $\widehat{\eta}, \widehat{\epsilon}$ are obtained, as will be seen in Section 3.2.

1.6. *Overview of the Paper.* In the remaining sections we rigorously establish the Bayesian framework presented above and report numerical experiments highlighting the performance of these estimators, possibilities for tuning and a practical application.

Sections:FEM shows how the formulation describedabove can be implemented using finite element numerical methods. In Section 3 we give an

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

implementation of our estimator and an application to actual data from simple molecular dynamics simulations arising in computational chemistry in Section 4 to make the point that our method is practically usable.

The mathematical detail is tackled in Sections 5 to 8. We proceed by first establishing existence and support of the prior measure in Section 5.

In order to approach the posterior measure we simply define a Gaussian measure with the mean and covariance suggested by (7) and then proceed to show that this definition satisfies all requirements of the posterior measure. In detail, we study the PDE relating prior and posterior means in Section 6 showing that the posterior mean is well-defined as the unique weak solution of the PDE (6). We continue by establishing existence and absolute continuity of posterior to prior measure in Section 7 and we verify (subject to a technical assumption) that the Radon-Nikodym derivative is in fact given by the likelihood (2) above.

It may be more natural to simply define the posterior measure as the product of the prior measure and the likelihood, however we have found it difficult to make our calculations rigorous using that approach.

We give a brief description of a simple method to estimate the local time from discrete high frequency observations in Section 8 but have to stress that this assumes adaptation of the spatial resolution to the observed time series.

Finally, in Section 9, we discuss natural extensions of this work.

1.7. *Literature Overview.* We present a very brief overview of the literature concerning non-parametric drift estimation which presently seems an understudied problem. Generally, we can classify the literature of non-parametric estimation for diffusion processes into a) classical/Bayesian, b) high/low frequency data. The first division refers to the way the unknown drift function is treated, i.e. as fixed unknown function to be estimated (classical) or as a random variable with a certain prior distribution (Bayesian). The second division refers to the type of data used in the estimation, that is whether it is assumed that effectively the continuous path is observed on a time horizon (high frequency) or the diffusion is only observed at discrete time points (low frequency). Classical non-parametric inference for high frequency data is a well studied problem, see for example [23]. Classical non-parametric inference with low frequency data is considerably harder, see for example the recent articles [9] and [3] and references therein. On the other hand, Bayesian non-parametric inference for diffusion processes is much less studied, and to the best of our knowledge this paper constitutes the first attempt in this direction. There is, however, some literature on Bayesian

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

non-parametric density estimation with striking similarities in the kind of prior and posterior measures used in [14]. Also, our work is motivated in part by known connections between between penalized nonparametric regression and Bayesian inference using Gaussian process priors, see [**?**].

This is in sharp contrast with the vast literature on Bayesian parametric inference for diffusion processes; for methodological work in this direction see for example [18], [17], [4], and for theoretical properties of Bayes estimators see for example [12] and [5].

2. Finite Element Implementation. In this section we briefly show an implementation of our estimator based on a finite element discretisation of the equations (6) and (7). For the sake of brevity we do not give details but follow standard practice in numerical analysis and provide pointers to the relevant literature. First though we shall briefly describe some essential mathematical preliminaries.

2.1. *Finite Elements.*

2.1.1. *Function Spaces.* Before defining the prior measure, we introduce notation for Sobolev spaces used in the definition as well as throughout the paper. We use the Sobolev spaces $H^s(U)$ (for $s \in \mathbb{N}$ and U an open bounded subset of some $\mathbb{R}^n$) of square integrable functions with s weak, square integrable derivatives with the inner product $(\cdot,\cdot)_{H^s}$ and induced norm $\|\cdot\|_{H^s}$. We denote the separable Hilbert subspace of functions with average zero by $\dot{H}^s(U)$. Furthermore, it is possible to impose periodic boundary conditions for Sobolev spaces of high enough order and we denote the Sobolev space of periodic functions on the interval (a,b) by $H^s_{\mathrm{per}}(a,b)$ for $s \geq 1$.

A convenient summary of Sobolev space fundamentals especially for the periodic case is given in [21] with a fuller treatment available from [11], whereas [2] contains a comprehensive treatment of this rich area. See also the neat treatment of weak solutions in [8].

We choose a Hilbert-subspace H^2_h of H^2 to discretize and implement the above estimator, in particular solving equations (6) and (7) respectively. H^2_h is just defined as the subset of H^2 consisting of cubic splines with continuous first derivatives, see Section 3.4.3 of [19] for details.

The fact that $H^2_h \subset H^2$ is referred to as using *conforming* finite elements and for H^2 in 1-dimension, piecewise cubic polynomials are the minimal requirement. Their regularity is sufficient by Theorem 5.2 in [7] and we employ these elements are recommended in [15] as well as [19].

We thus represent functions $u \in H^2_h$ by piecewise cubic polynomials on the B elements discretizing the interval $[0, 2\pi]$. The operators I, Δ and Δ^2

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

are then discretised as the matrix encoding the action of the associated quadratic form on H_h^2 using standard finite element theory, e.g. from [7]. Local time is discretised using a histogram with B bins and piecewise constant (P^0) functions leading to a natural discretisation of the linear form $\phi(u) = \int_0^{2\pi} u'(a)L(a)da$. The trilinear form $\psi(u, L, v) = \int_0^{2\pi} u(a)L(a)v(a)da$ is also straightforward to discretize, somewhat akin to standard mixed finite element discretizations of the trilinear form in the Navier Stokes equation, see e.g. Chapter III.5 in [7] in the context of the Stokes equation.

Exploiting the coercivity shown in Lemma 1 in the appendix, Section 10.2, it is then possible to show that the solutions to the discretised system are only $\mathcal{O}(h)$ away in H^2-norm from the true solution, using the general version of Céa's Lemma from [15]. This type of result is often stated as the error of approximation being bounded above by the error of interpolation:

THEOREM 1. *Let u be the true solution of (6) with $\eta = \epsilon = 1$, $k = 2$ and periodic boundary conditions (formally described in (30)) and let u_h be the spline function in $\mathcal{H}_h^2$ corresponding to the solution U of the discretised problem*

$$S_2U + MU + M_LU = M(W + \tilde{\chi}_h(\cdot; X_0, X_T)) + CL_{T,h} \tag{14}$$

where S_2, M and M_L are matrices arising from suitable finite element discretizations of the differential operator Δ^2 and the multiplication operators I and $L_T \cdot I$ respectively, then we have the bound

$$\|u - u_h\|_{H^2} \leq \frac{C}{\alpha} \inf_{v_h \in \mathcal{H}_h^2} \|u - v_h\|_{H^2}, \tag{15}$$

where the constants α and C are the constants of coercivity and of continuity as in Lemma 1.

To bound the error of interpolation, we refer to [19], Section 3.4.3.

3. Experimental results.

3.1. *Setup.* As a concrete example we study the SDE

$$dx = \Big(-\sin(x) + 3\cos^2(x)\sin(x)\Big)\,dt + dB, \quad x(0) = \frac{2}{3}\pi \tag{16}$$

The Exact Algorithm 1 from [1] is used to generate a sample path for final time $T = 2000$. The resolution timestep is $\delta t = 0.001$ and only the samples at time points $\{j\delta t\}_{j=1}^{2\cdot 10^6}$ are used in the computation of the histogram for

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

the local time (i.e. potential intermediate skeleton points are discarded). For the portions of that path obtained by truncating to final times $T \in \{250, 500, 1000, 2000\}$ respectively we compute the histogram on a uniform $B = 50$-bin discretisation of the interval $[0, 2\pi]$.

The prior measure specified is a Gaussian with zero mean and covariance operator $(\eta\Delta^2 + I)^{-1}$ for $\eta \in \{\frac{1}{2000}, \frac{1}{200}, \frac{1}{2}\}$.

Fitted posterior means (i.e. solutions of (6) are given in Figure 1. This figure also gives the maximum likelihood solution

$$L_T\hat{b} = \frac{1}{2}L'_T$$

to which we refer as local time solution. Note that this is not quite the same as just the logarithm of the discretised local time since an interpolation to $\mathcal{H}^2_h$ takes place.

Looking at Figure 1 it is clear that as the final time increases, the posterior mean converges to the true drift. Since the drift is smooth ($\mathcal{C}^\infty$) convergence is unaffected by the specified prior. The prefactor η in the specification of the prior is critical - the quality of approximation of the true drift for medium range final times greatly depends on tuning this parameter as can be inferred contrasting Figures 2 and 3 to Figure 1.

3.2. *Optimal Smoothness Parameter.* Further to Subsection 1.5, we use the expression (13) to derive optimal values for the parameter η and highlight potential problems with the empirical Bayes approach in this context.

Testing this for a single fixed sample path ($T = 800$), in the case $k = 1, \epsilon = 1$, the resulting optimal values for η as a function of number of mesh elements B are given in Figure 4 together with the behaviour of the normalizing constant. The parameter ϵ has been studied, too, optimizing over both parameters. The results are only slightly different, so in order to simplify the presentation, we restrict attention to the case of fixed $\epsilon = 1$.

In the case of sample paths of this length, the procedure is remarkably robust wrt. changes of the number of mesh elements B, in spite of the fact that the prior measures for different values of η are mutually singular! From inspection of Figure 4 it seems that the normalization constant calculated in (13) is continuous in η. Also, inspection of the posterior means as given in Figure 5 shows that while choosing the correct order of magnitude is important, the exact choice of η is not critical and that the optimal choice in terms of normalization constant is not too far from optimal in terms of the L^2-distance of the posterior mean from the true drift function, where the latter distance would be unavailable in practical applications, of course.

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

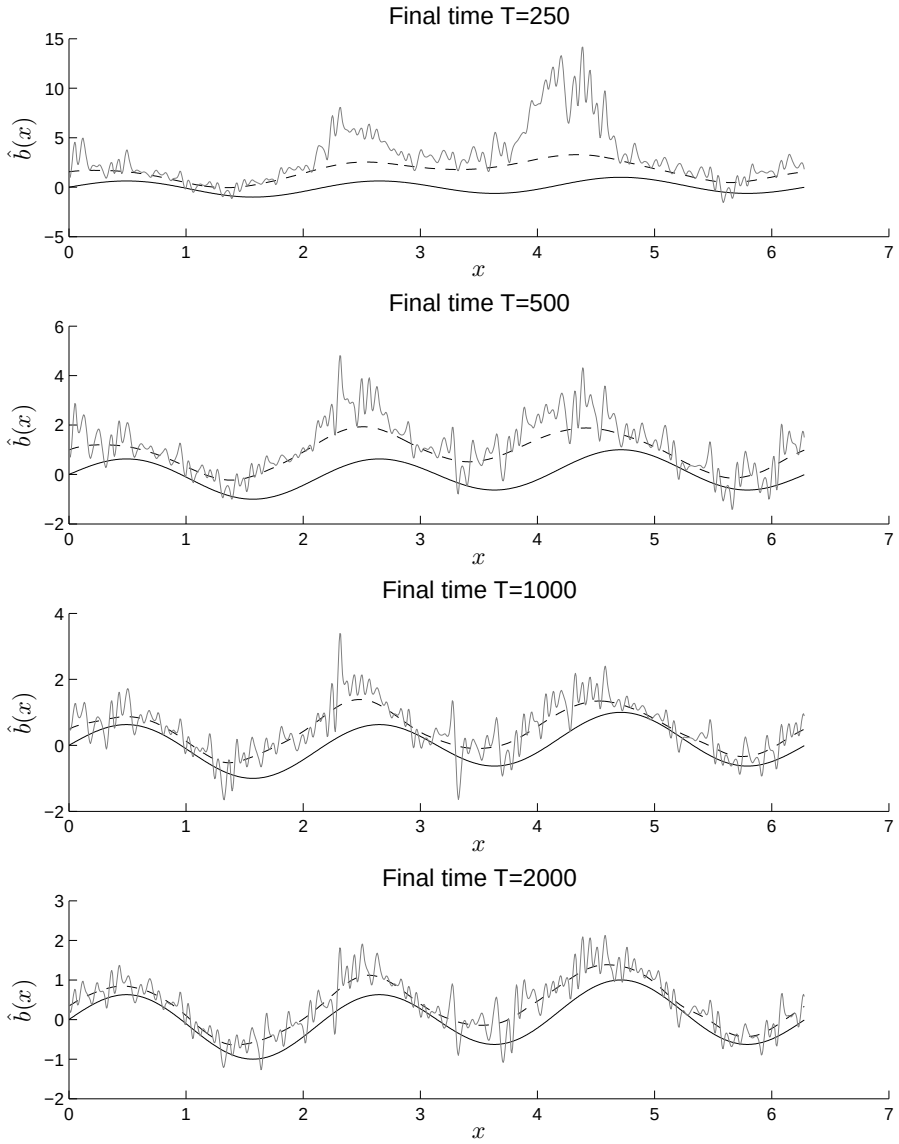

FIG 1. *Posterior Means for increasing final time T, $\eta = \frac{1}{200}$*
Solid Line: True Drift, Dashed Line: Posterior Mean, Grey Line: Local Time Solution

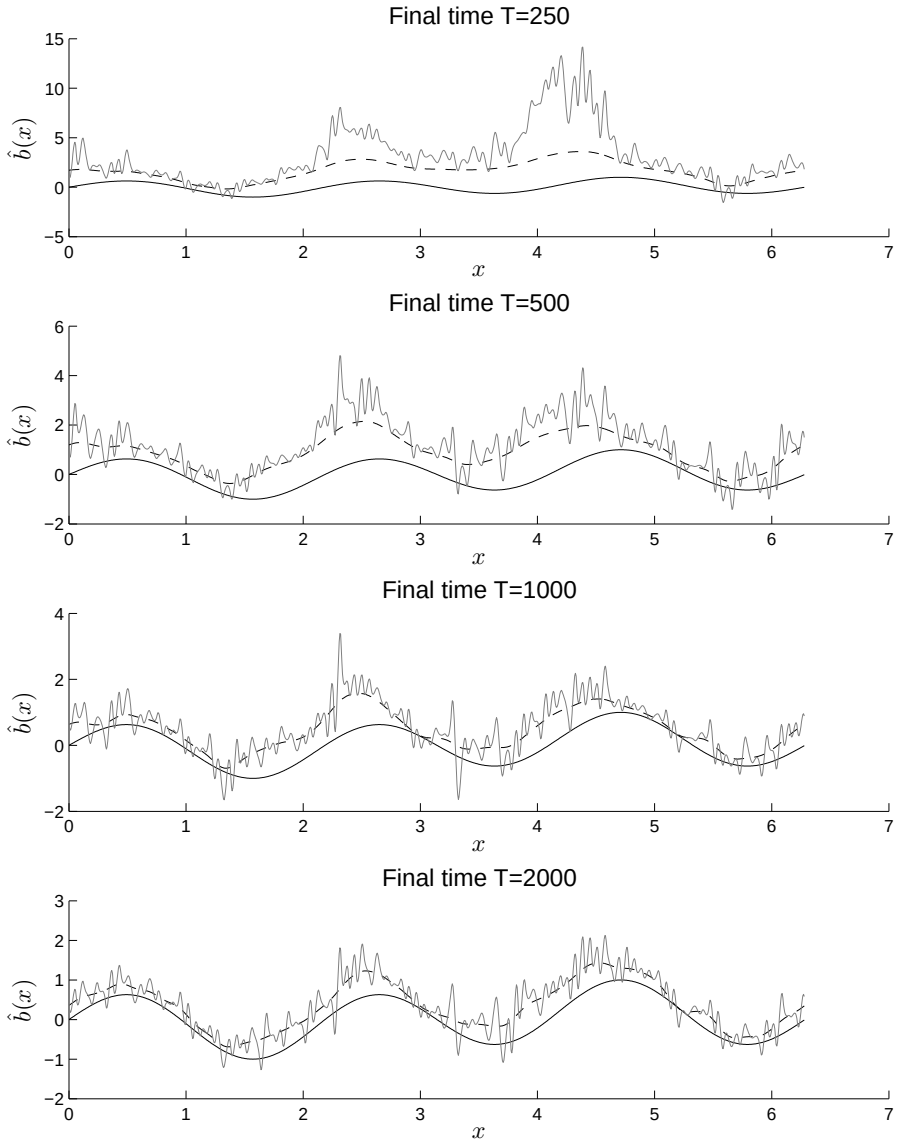

FIG 2. *Posterior means for increasing final time T, $\eta = \frac{1}{2000}$ Solid Line: True Drift, Dashed Line: Posterior Mean, Grey Line: Local Time Solution*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

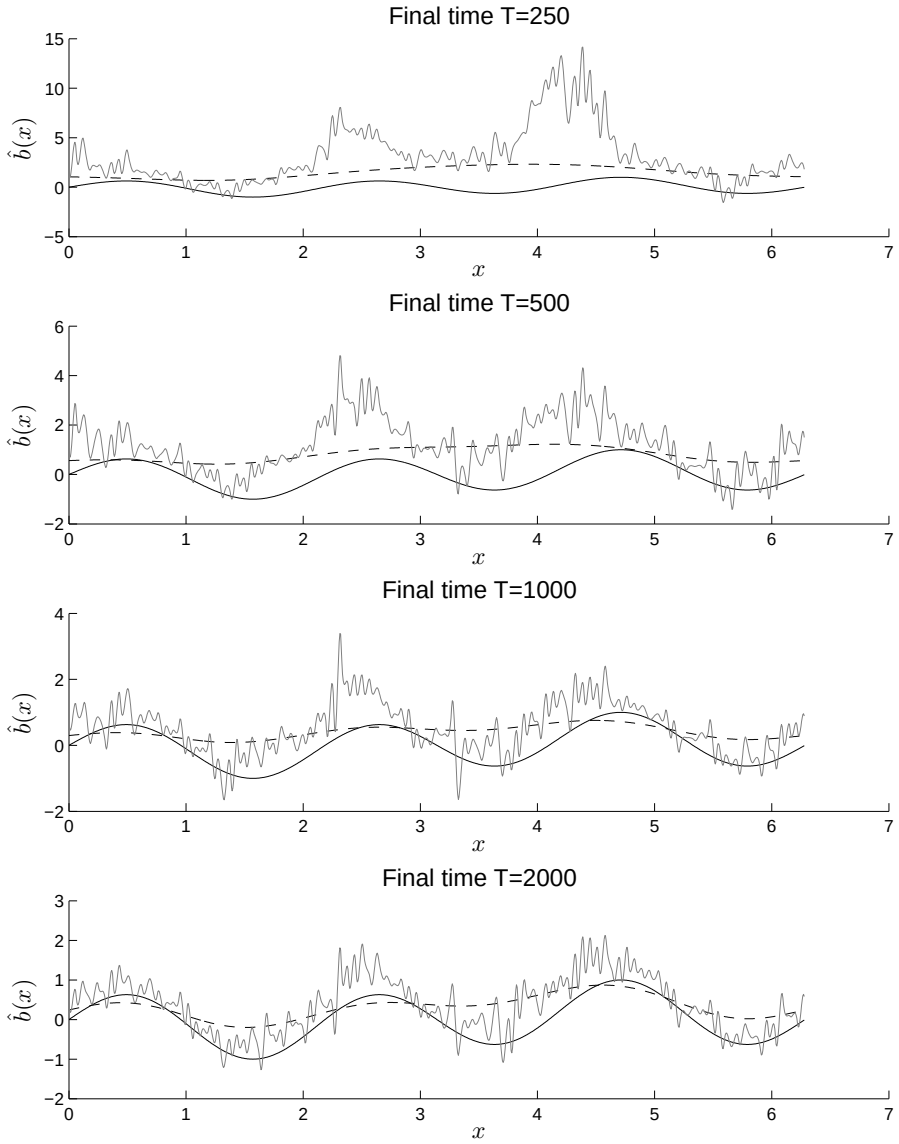

FIG 3. *Posterior means for increasing final time T, $\eta = \frac{1}{2}$*
Solid Line: True Drift, Dashed Line: Posterior Mean, Grey Line: Local Time Solution

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

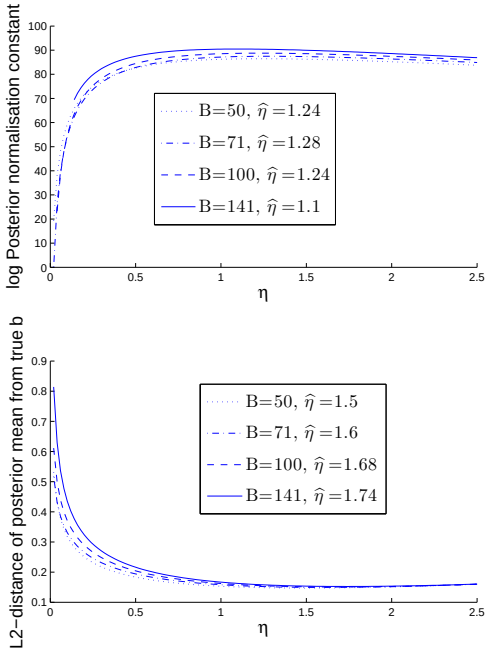

FIG 4. *Log-normalization constants and L^2-distances vs. hyperparameter η*

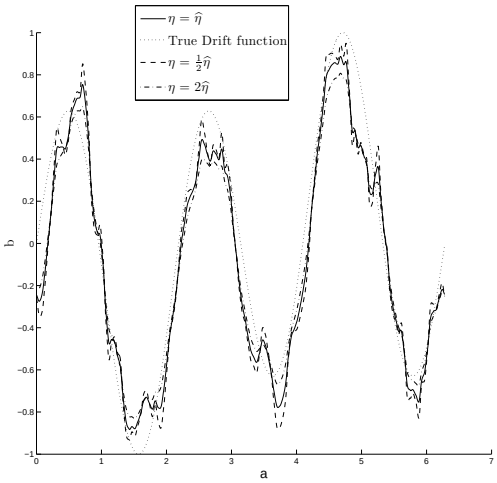

FIG 5. *Posterior mean drifts for optimal and some suboptimal choices of η*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

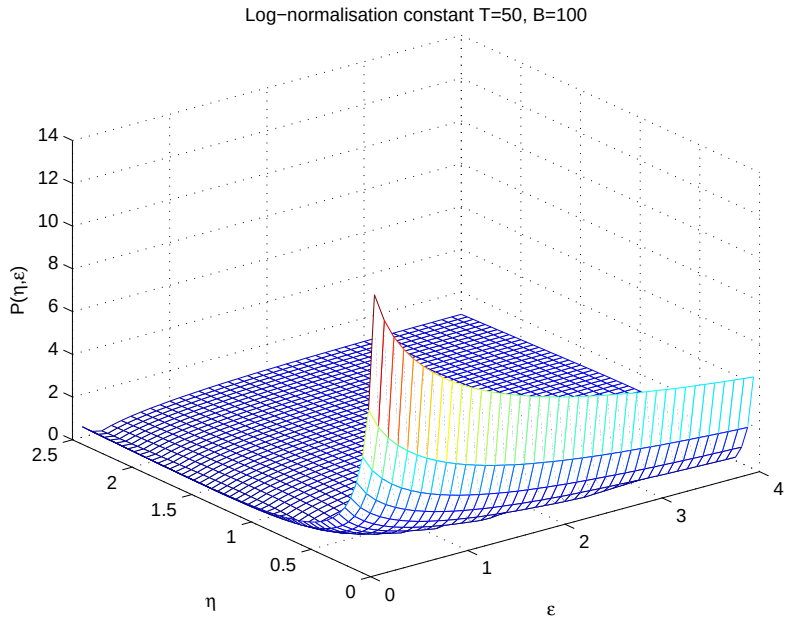

FIG 6. *Log-normalization constant vs. η and ϵ – short path: $T = 50$*

For shorter final times e.g. in the case $T = 50$, especially when the local time is zero on some interval in the circle, the procedure sometimes yields optimal hyperparameters $\widehat{\eta} = 0, \widehat{\epsilon} = 0$, which is of course unacceptable since neither prior nor posterior measures exist in this case; see Figure 6. We have investigated the phenomenon numerically and it seems that it is a manifestation of a problem at the analytical level, i.e. it does not seem to be an artifact of discretisation. However, an analytical treatment of the problem seems difficult, and since it is not unusual for empirical Bayes to yield optimal values on the boundary of the parameter space, we do not investigate the problem further.

3.3. *Order of Covariance Operator.* Another aspect to study is the order of the covariance operator used in specifying the prior. Resulting posterior means for the force functions are given in figure 7 for the local time solution, the second-order covariance operator (i.e. $(\eta\Delta + I)^{-1}$) abbreviated to S1 and the fourth-order covariance operator (i.e. $(\eta\Delta^2 + I)^{-1}$) abbreviated to S2. The parameters are $T = 1000$ and $\eta = 1$.

Some actual samples from the posterior measure are given for the force

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

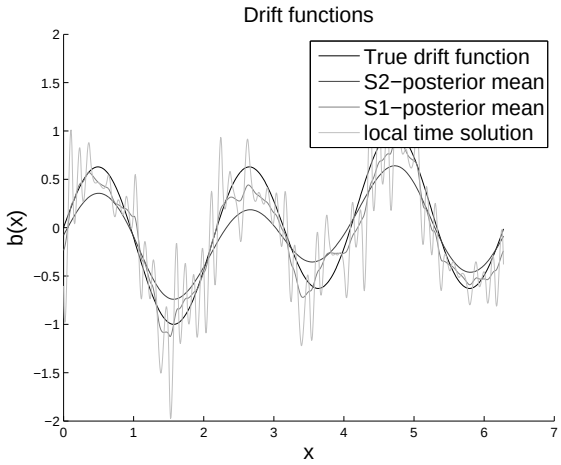

FIG 7. *Posterior means using S2, S1 priors and local time solution*

functions in the S1 and S2 cases in figure 8 for $\eta = \frac{1}{10}$ but $B = 500$ finite elements to render the regularity of the sample paths more visible.

Gaussian Boundary Conditions. Further to the formal calculation presented in Subsection 1.4, we implement a numerical solution of (12) using the same finite element implementation as above and implement the boundary conditions using the functional (9). For the sake of contiguous presentation, we consider the process

$$dx = -2\tanh(x-\pi)dt + dB, \quad x(0) = \pi \tag{17}$$

where the scaling factor 2 is chosen to make the process spend most of its time in $[0, 2\pi]$, so that it makes sense to choose $y = 0$ and $z = 2\pi$ in Subsection 1.4. We use a smoothing constant $\eta = 5$ and a boundary point prior variance of $\sigma^2 = 10$, as used in (12). Again, we use a sampler based on Exact Algorithm 1 from [1] with similarly small time increment δt as above. With these conventions, we obtain the plots in Figures 9, 10 and 11. Since our posterior measures on the space of force functions leave it unclear what happens to the process once it leaves $[0, 2\pi]$, we cannot decide whether ergodicity is maintained.

3.4. *Rates of Posterior Contraction.* Convergence is studied for the choice $\eta = 2 \cdot 10^{-4}$, $\epsilon = 0$. This particularly small choice enables convenient investigation of issues connected with finite numerical resolution. Note that for

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

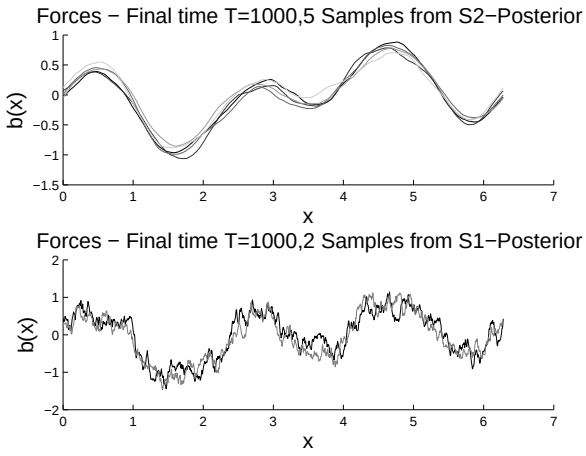

FIG 8. *Draws from posterior drifts, S1 and S2 cases*

this value of η, the difference between the MLE (in its finite element implementation) and the Bayesian posterior mean is nearly invisible in any of the plots given before.

In considering how the posterior distribution converges as $T \to \infty$ it is clear that marginalization of some form will be needed to make the convergence of measures on the infinite dimensional space $H^1(0, 2\pi)$ amenable to numerical investigation. Since we are interested in the behaviour of the proposed method under mesh refinement, merely considering the finite dimensional finite element discretisation as the marginalization is unhelpful. While many more sweeping marginalizations are conceivable, we will concentrate on three marginalizations of the posterior measure μ:

- **Point evaluation:** Consider $\tau(u) = u(0.38\pi)$
- **Integration:** Integrate against a test function: $\tau(u) = \int_0^{2\pi} u(a) \sin(a) da$
- **Norms:** Appropriate Hilbert Space norms: $\tau(u) = \|u\|_{L^2}$ or $\tau(u) = \|u\|_{H^\gamma}$, $\gamma \in (0, 1)$.

Note that point evaluation and integration are linear operations, so that the posterior is expected to be Gaussian, whereas taking norms will result in distributions more akin to χ^2.

3.4.1. *Point Evaluation.* We look at the point value of the posterior mean $\hat{b}$ at $x = 0.42\pi$. The resulting posterior densities for $\hat{b}(0.42\pi)$ display an approximately Gaussian shape and the mean deviations and their variances are given in Figure 12. It is clear that the functional which picks

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

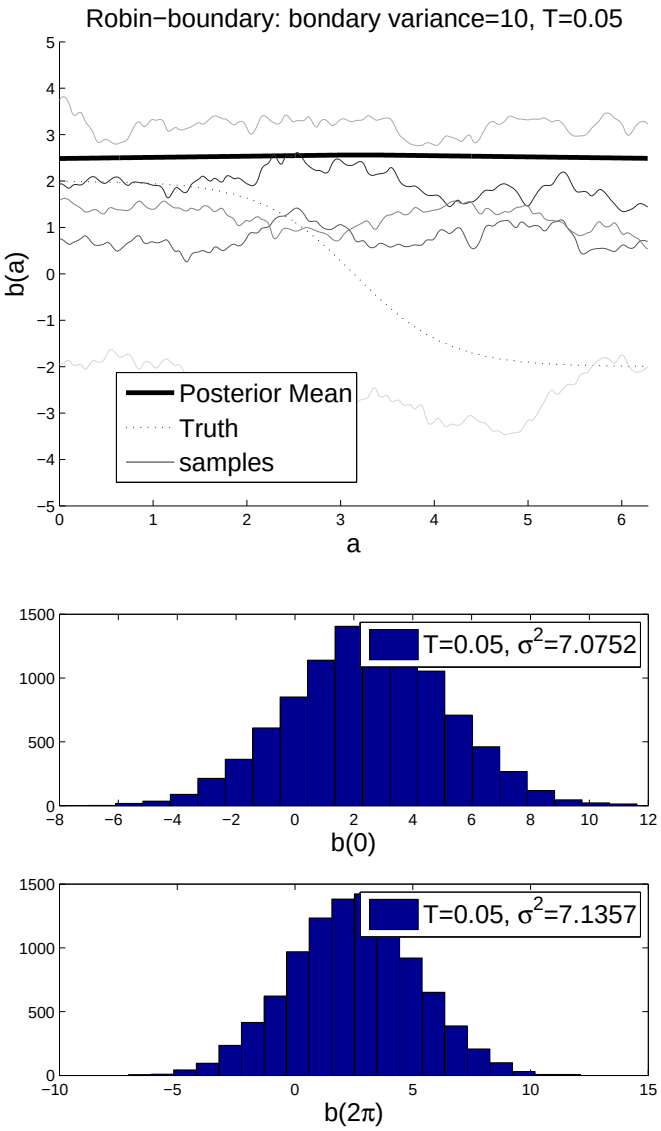

FIG 9. *Posterior Forces using Gaussian BCs, T=0.05*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

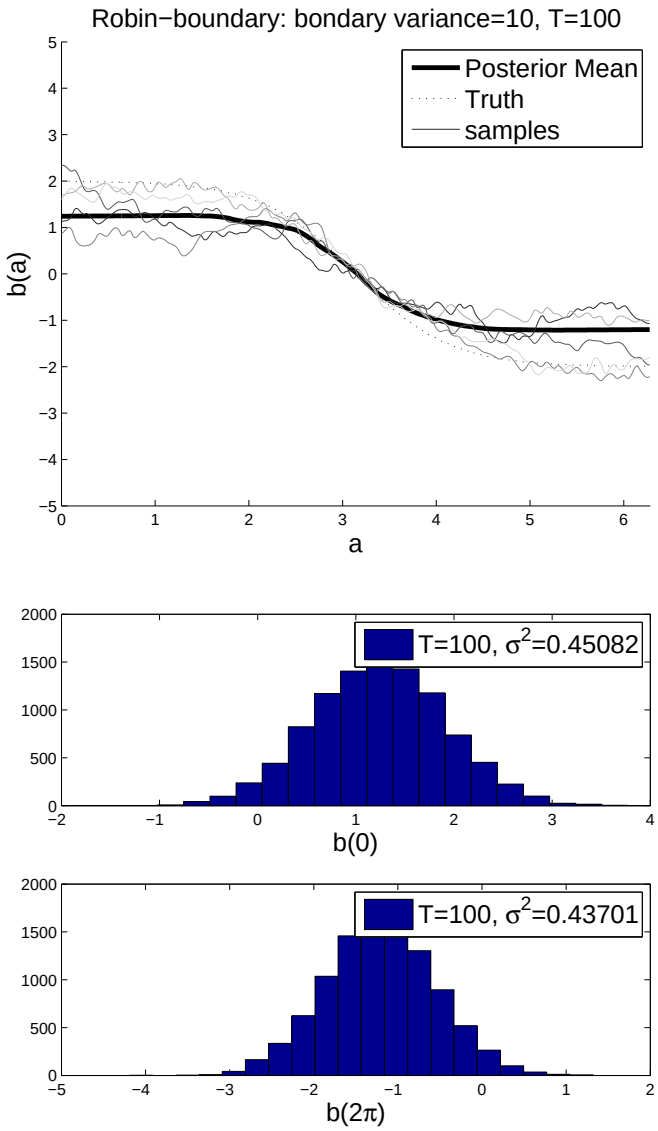

FIG 10. *Posterior Forces using Gaussian BCs, T=100*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

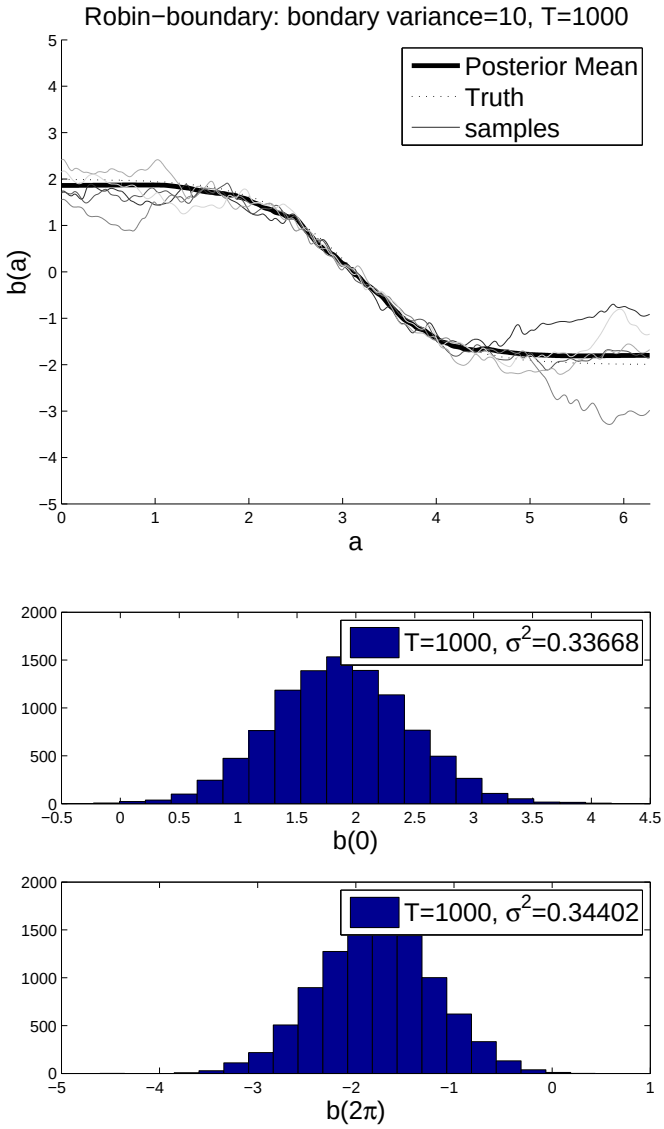

FIG 11. *Posterior Forces using Gaussian BCs, T=1000*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

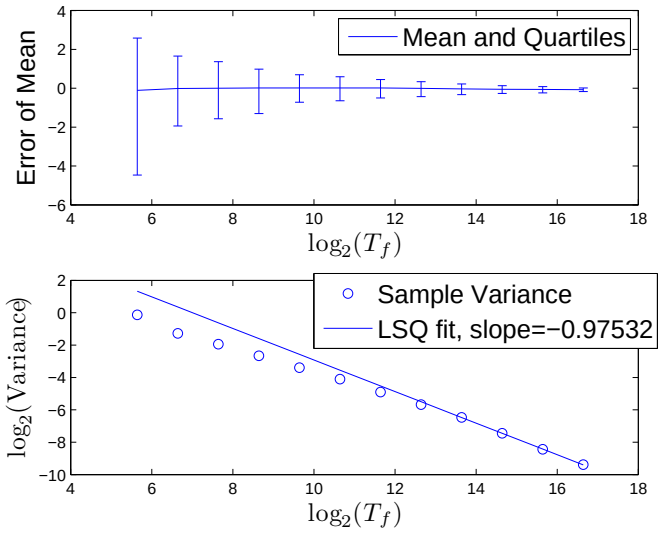

FIG 12. *Convergence of deviations of the posterior mean for* $b(0.42\pi)$

the point value of $\hat{b}$ has a strong high-frequency component when subjected to Fourier transform. Looking at an eigenvalue analysis of the PDE (6) this would suggest that the final time needed to see the CLT-like decay of variance like $\mathcal{O}(1/T)$ is very large and linked to the numerical resolution. In fact, one can perform a principal component analysis of the variance to find that most of the randomness resides in the high-frequency components of the right hand side of the discretised PDE.

To investigate this issue more closely, four different finite element resolutions, $N \in \{35, 50, 71, 100\}$ have been used to observe the decay of variances of the posterior marginalized means. Consulting the right hand side plot in Figure 13, where point evaluation $b(0.42\pi)$ has been used to marginalize as before, it is apparent that a CLT-like decay of $\mathcal{O}(1/T)$ is only achieved for the lower resolutions. Furthermore, the $1/T$-asymptotics take over at increasingly larger final times T with increasing number of mesh elements N.

3.4.2. *Integration against Test Function.* On the other hand, when the procedure to marginalize the posterior means is to numerically integrate against the function $\sin(x)$, i.e. to consider the random variables $\zeta = \int_0^{2\pi} \sin(a)\hat{b}(a)da$, then it is apparent from the top plot in Figure 13 that the $1/T$-asymptotic is much easier to observe and that its onset is essentially

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

independent of the chosen numerical resolution, N.

In conclusion, a CLT-like decay of variance is observed numerically for the posterior means as expected when evaluation is against a smooth functional; and this asymptotic is found to be robust against numerical resolution. For point evaluation, however, absence of the CLT-like decay of variance for the continuous PDE where the spectrum of the differential operator Δ^2 is *not* bounded (as opposed to its discretisation where it always is bounded) is supported by the numerical observations.

It would be interesting to investigate whether the intermediate slopes for the decay of variance observed in Figure 13 – e.g. for $N = 100$ the slope of approximately -0.4 – can be accounted for using the above-mentioned Fourier picture. While this is not central to the newly proposed Bayesian nonparametric estimation procedure, it might have mathematical merit.

3.4.3. *Norms.* Similar experiments can be performed when norms are used to examine convergence, although the resulting distribution will of course not be Gaussian but rather more like χ^2. Some example distributions can be seen in Figure 14.

Again, one can consider the mean and variance of posterior errors which is carried out in Figure 15. Note that convergence of the means is only observed in the L^2 case with convergence in the H^1 case only occurring beyond $\log_2(T_f) \geq 12$ which is the regime where the finite mesh size is overcome by large final time, so that no convergence would be expected there in the case of increased resolution.

4. Application Example: Molecular Dynamics. In this section we give a brief application of the proposed non-parametric drift estimator to a toy example from Molecular dynamics. We start by explaining the origin of the data, apply the fitting algorithm to it and then compare the results to previous fits found in the literature.

The data used for this fitting example are generated using a molecular dynamics (MD) simulation for a single molecule of Butane. In order to avoid explicit computations for solvent molecules, several *ad hoc* approximate algorithms have been developed in molecular dynamics. One of the more sweeping approximations that is nonetheless fairly popular, at least as long as electrostatic effects of the solvent can be neglected or treated otherwise, is Langevin dynamics. Here, the time evolution of the Cartesian coordinates $x \in \mathbb{R}^{12}$ of the four extended atoms of Butane (see Figure 16)

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

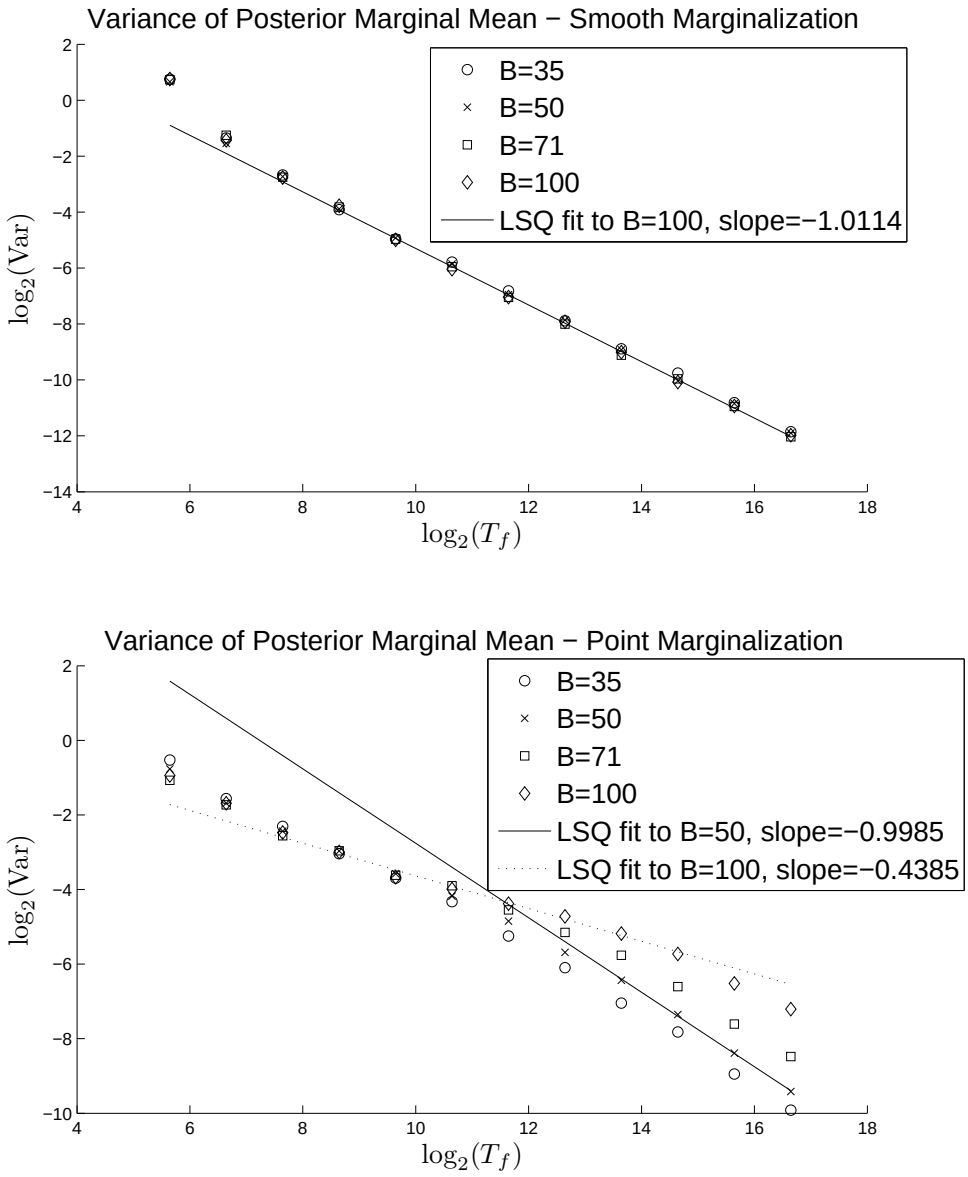

FIG 13. *Convergence of variances of the marginalized posterior means*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

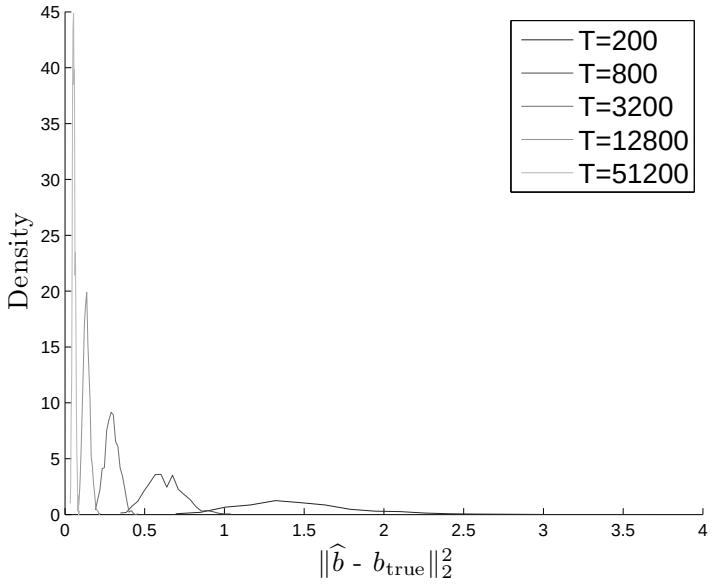

FIG 14. *Distribution of L^2-marginals for various final times T*

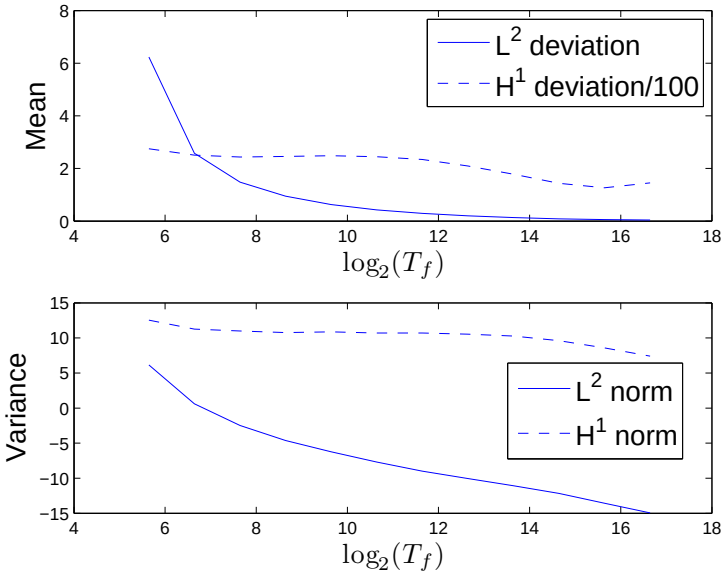

FIG 15. *Means and variances of H^1 and L^2-norm marginalizations*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

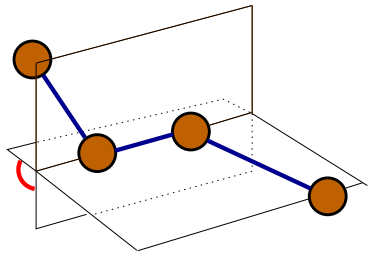

FIG 16. *Sketch of Dihedral Angle*

is simulated using a damped-driven Hamiltonian system:

$$\ddot{x} + \gamma\dot{x} - \nabla V(x) = \sqrt{\frac{2\gamma}{\beta}}\dot{W}. \tag{18}$$

This is a second order hypoelliptic diffusion process driven by standard Brownian motion W, $\gamma > 0$ is the friction coefficient and $\beta > 0$ plays the role of inverse temperature. Details of the force field and potential V used here can be found in [10] and an overview of the use of Langevin dynamics in Molecular Dynamics simulations can be found in [22].

From a chemical point of view interest is focused on the dihedral angle ω, which is the angle between the two planes in $\mathbb{R}^3$ formed by atoms $1, 2, 3$ and atoms $2, 3, 4$ respectively; see the sketch in Figure 16. Conformational change is manifest in this angle, and the Cartesian coordinates themselves are of little direct chemical interest. Hence it is natural to try and describe the stochastic dynamics of the dihedral angle in a self-contained fashion.

One MD run is produced using a timestep of $\Delta t = 0.1$fs (Throughout this section, we use the time unit femtosecond abbreviated to fs. Note that $1fs = 10^{-15}s$.) and a Verlet variant (see p.435 in [22]) covering a total time of $T = 4 \cdot 10^{-9}$s (4 nanoseconds). A section of the path of the dihedral angle as a function of time can be seen on the left of Figure 17; the corresponding histogram for the whole of the path is depicted to the right of that figure.

It should be stressed that the Itô process governing the behaviour of the dihedral angle ω is *not* of the form (1), in particular, it will have a non-constant diffusivity and be hypoelliptic of second order, so its regularity will be C^1.

4.1. *Fitting.* We aim to non-parametrically estimate the drift function $b(\cdot)$ in (1 to the trajectory shown in Figure 17. There is an arbitrary choice of timestep to make and we make this such that the resulting quadratic

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

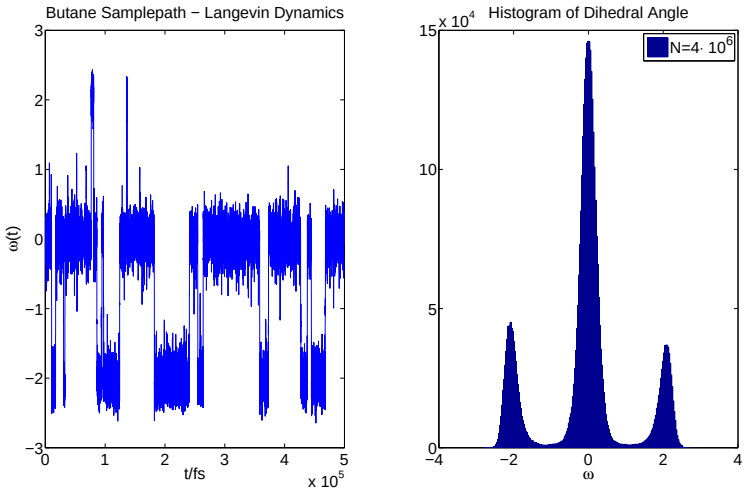

FIG 17. *MD Samplepath Butane.*
Left: First 500ps of sample path, Right: Histogram of whole sample path

variation

$$\langle X \rangle = \sum_{i=0}^{n-1} (X_{i+1} - X_i)^2 \tag{19}$$

corresponds to $\langle X \rangle = T$ so that we have apparent diffusivity 1. To understand the next preprocessing step of subsampling, consider Figure 18 where we have plotted the apparent quadratic variation as a function of subsampling factor k, i.e. we computed (19) using only every kth step of the Butane data. From this Figure, it is clear that the region $k \in \{1, \ldots, 100\}$ is not usable since this shows a step $\sqrt{k}$-law increase of apparent diffusivity – this mirrors the fact that the path actually has quadratic variation zero since it is of smoothness $\mathcal{C}1$. The region $k \in \{100, \ldots, 500\}$ also is not very usable since the apparent diffusivity varies strongly, so that the process is not adequately described by (1) on this timescale, either. On the timescale obtained for subsampling factors $k \in \{1000, \ldots, 5000\}$ the apparent diffusivity varies little so that we settle for a subsampling factor $k = 1000$ taking the observed apparent diffusivities form Figure 18 as indication that the dihedral angle process behaves qualitatively like a first order diffusion on those timescales. The attributed final time to make the diffusivity one is $T = 1127$.

Now, using a Gaussian prior measure with mean zero and fourth order covariance operator with periodic boundary conditions and the parameter

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

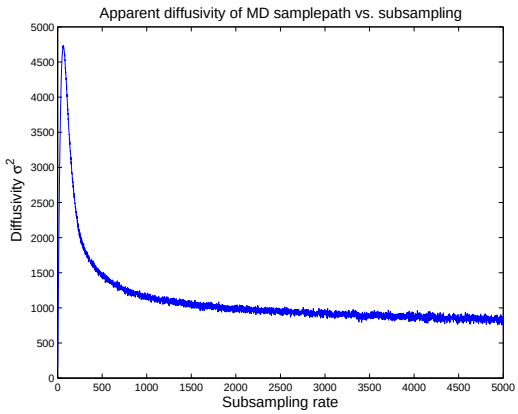

FIG 18. *Apparent Diffusivity versus subsampling factor k*

$\eta = 0.02$ we perform non-parametric estimation of the drift function in (1) for the Butane data set described above. The resulting posterior mean is given in Figure 19 where the solid line indicates the posterior mean and the shaded regions display one standard deviation credibility bands derived from the posterior measure. In the case of this dataset, the posterior normalization constant used in Subsection 3.2 yields an optimal hyperparameter $\widehat{\eta} = 0$, see Figure 20.

5. Existence and support of the prior measure.

5.0.1. *Existence.* We show in Appendix 10.1 that the covariance operator $(\Delta^2 + I)^{-1}$ for the Gaussian measure

$$\mu_0 = \mathcal{N}\left(0, (\Delta^2 + I)^{-1}\right) \tag{20}$$

on $L^2(0, 2\pi)$ exists, and that it is symmetric, non-negative and trace-class and can thus use Theorem 2.3.1 from [6] to establish the existence of a corresponding Gaussian measure.

5.0.2. *Support.* The Cameron-Martin-space for the measure μ_0 is $\mathcal{H} = H^2_{per}(0, 2\pi)$ equipped with the norm $\|f\|^2_{\mathcal{H}} = \|\Delta f\|^2_{L^2} + \|f\|^2_{L^2}$. So, we use a series representation of the random variable $X \sim \mu_0$ as follows:

$$X = \sum_{k=-\infty}^{\infty} \xi_k(\omega) e_k \tag{21}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

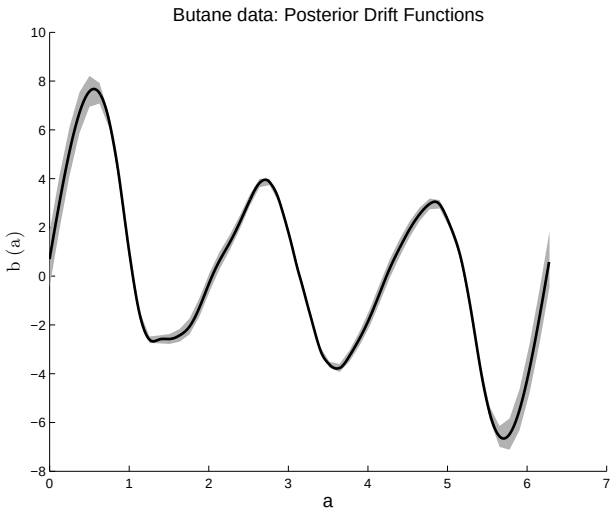

FIG 19. *Posterior mean drift function for Butane data*

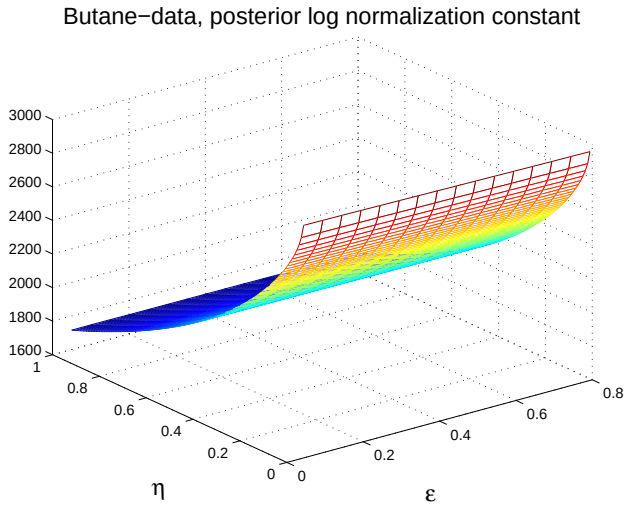

FIG 20. *Posterior log-normalization constant for Butane data*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

where the e_k are the following orthonormal basis of $\mathcal{H}$:

$$e_k(x) = \begin{cases} \frac{1}{\sqrt{\pi(k^4+1)}} \cos(kx) & k > 0 \\ \frac{1}{\sqrt{2\pi}} & k = 0 \\ \frac{1}{\sqrt{\pi(k^4+1)}} \sin(kx) & k < 0 \end{cases}, \tag{22}$$

and the ξ_k are iid. standard normal random variables. Theorem 3.5.1 in [6] guarantees that the representation (21) converges almost everywhere with respect to μ_0. To find out whether $X \in H_{per}^{\frac{3}{2}-\epsilon}$ for $0 < \epsilon < \frac{3}{2}$, compute the appropriate seminorm as follows:

$$\begin{aligned} |X|^2_{H^{\frac{3}{2}-\epsilon}} &= \sum_{k=-\infty}^{\infty} \xi_k^2 |k|^{3-2\epsilon} \|e_k\|^2_{L2} \\ &= \sum_{k=-\infty}^{\infty} \xi_k^2 |k|^{3-2\epsilon} \frac{1}{1+k^4} \\ &\leq \sum_{k=-\infty}^{\infty} \xi_k^2 |k|^{-1-2\epsilon}. \end{aligned}$$

The latter sum converges almost surely for any $\frac{3}{2} > \epsilon > 0$. Therefore, a draw from μ_0 will almost surely be contained in $H^{\frac{3}{2}-\epsilon}$ for any $\epsilon > 0$. Thus, we have found a Hilbert support of μ_0.

6. Analysis of the PDE for the Posterior Mean. In this section we present a brief analysis of the PDE for the posterior mean, (6), establishing coercivity, continuity, symmetry and positivity of the differential operator under mild conditions on the local time L_T, and consequently existence and uniqueness of solutions.

Stability against small, admissible perturbation of L_T is also established and enables error estimates for the whole approximation process starting from the $L2$-error in the local time L_T up to the *numerical* solution u.

6.1. *Analytical Setup.* We write the PDE (6) with the choice $\mathcal{A} = \Delta^2$ for the prior as

$$\left(\Delta^2 + I + L_T\right) u = \frac{1}{2} L_T' + W + \tilde{\chi}(\cdot; X_0, X_T), \quad u \in H2_{\text{per}}([0, 2\pi]) \tag{23}$$

(24)

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

Here the PDE is solved in the weak H^2-sense, i.e. we understand (23) to mean that we seek a $u \in H^2_{\text{per}}([0,2\pi])$ such that

$$\int_0^{2\pi} \Delta v(a)\Delta u(a) + (1+L_T(a))u(a)v(a)da =$$
$$\int_0^{2\pi} -\frac{1}{2}v'(a)L(a) + (W+\tilde{\chi}(a;X_0,X_T))v(a)da \quad \forall v \in H^2_{\text{per}}.$$

See [11] for details. We make use of the abbreviation f introduced in (1.3).

6.2. *Coercivity.* In order to move to a rigorous weak formulation of the PDE problem (23) we write it using the quadratic form

$$a_G(u,v) = \int_0^{2\pi} \Delta u(a)\Delta v(a) + u(a)G(a)v(a)da, \quad u,v \in H^2_{\text{per}}([0,2\pi]) \tag{25}$$

Assuming that the function G is continuous, periodic, non-negative and not identically zero we obtain coercivity of this bi linear form:

LEMMA 1. *If function G is continuous, periodic, non-negative and not identically zero on $[0,2\pi]$, then the form a_G, defined in (25), is a continuous, coercive, symmetric bi linear form, i.e. there are constants $\alpha, C \in \mathbb{R}_+$ which may depend on G but not on u,v such that the following relations hold:*

$$\alpha\|u\|^2_{H^2} \le a_G(u,u) \quad \forall u \in H^2_{\text{per}}([0,2\pi]) \tag{26}$$
$$a_G(u,v) \le C\|u\|_{H^2}\|v\|_{H^2} \quad \forall u,v \in H^2_{\text{per}}([0,2\pi]) \tag{27}$$
$$a_G(u,v) = a_G(v,u) \quad \forall u,v \in H^2_{\text{per}}([0,2\pi]) \tag{28}$$

The proof of this lemma, in particular the proof of coercivity is slightly technical due to the differential operator involved being of fourth order, so that we present the proof of this lemma in Appendix 10.2.

Note that the local time L_T almost surely satisfies the hypotheses on the function G needed in Lemma 1, *a fortiori*, $L_T + 1$ also satisfies these hypotheses. For future use, we note down the quadratic form as used in in this case, i.e. for $G = I + L_T$:

$$a(u,v) = \int_0^{2\pi} \Delta u(a)\Delta v(a) + u(a)(1+L_T(a))v(a)da, \quad u,v \in H^2_{\text{per}}([0,2\pi]) \tag{29}$$

6.3. *Existence and Uniqueness.* Now that coercivity and continuity are established, existence and uniqueness of solutions for the PDE (23) can be inferred in the usual way from the Lax-Milgram Lemma. For the particular case at hand, though, the fact that the quadratic form (25) is symmetric enables the use of the Riesz representation theorem as follows.

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

THEOREM 2. *Let $L_T \in \mathcal{C}([0,2\pi])$ be continuous and periodic and not identically zero. Then the PDE*

$$\Delta^2 u + (I + L_T)u = \frac{1}{2}L_T' + W + \tilde{\chi}(\cdot; X_0, X_T) \tag{30}$$

has a unique solution $u \in H^2_{\text{per}}([0,2\pi])$.

PROOF. To show this, note that f from (1.3) is at least of regularity H^{-1} since $L_T \in \mathcal{C}_{\text{per}}([0,2\pi]) \subset L^2([0,2\pi])$ and so its derivative will be in H^{-1}: $L_T' \in H^{-1}([0,2\pi])$, *a fortiori* we have $L_T' \in H^{-2}([0,2\pi])$. Furthermore $\tilde{\chi}(;X_0, X_T) \in L^2([0,2\pi])$ and of course the same holds for the constant function W, so the same argument yields $f \in H^{-2}([0,2\pi])$. We have shown in the above lemma 1 that the bi linear form associated with the partial differential operator $\Delta^2 + I + L_T$ is symmetric, continuous and coercive and hence defines an inner product on $H^2_{\text{per}}([0,2\pi])$. Therefore, by the Riesz representation theorem, there exists a unique $u \in H^2_{\text{per}}([0,2\pi])$ such that

$$\int_0^{2\pi} \Delta u \Delta v + u(1 + L_T)v dx = \int_0^{2\pi} v f dx \quad \forall v \in H^2_{\text{per}}([0,2\pi]).$$

This u is the unique weak solution to the PDE (30) in the statement of the theorem. □

6.4. *Continuous Dependence on L_T.* In this subsection we show robustness of the posterior mean against small errors in the local time L_T that satisfy some positivity assumptions, in particular that the perturbed L_T stays positive enough in a sense to be made precise. This result is not entirely obvious since the map from the local time L_T to the posterior mean u is *nonlinear*.

We establish the result by analyzing the PDE (23) for the perturbed local time $\tilde{L}$:

$$\left(\Delta^2 + I + \tilde{L}\right)\tilde{u} = \frac{1}{2}\tilde{L}' + \tilde{\chi}(a, X_0, X_T) + W, \quad \tilde{u} \in H^2_{\text{per}}([0,2\pi]) \tag{31}$$

This PDE can be rewritten in terms of the unperturbed local time L_T and the perturbation $\tilde{L} - L_T$:

$$\left(\Delta^2 + I + L\right)u + \left(\Delta^2 + I + L\right)(\tilde{u} - u) + (\tilde{L} - L)\tilde{u} = \frac{1}{2}\tilde{L}' + \tilde{\chi}(\cdot; X_0, X_T) + W$$

Using that u satisfied the PDE (23), we can subtract this PDE from the previous one to obtain:

$$\left(\Delta^2 + I + \tilde{L}\right)(\tilde{u} - u) = \frac{1}{2}\left(\tilde{L}' - L'\right) - \tilde{u}\left(\tilde{L} - L\right) \tag{32}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

Note that we have assumed for the moment that W and $\tilde{\chi}(\cdot; X_0, X_T)$ are unperturbed which certainly for discrete time observations of initial and final condition is safe for $\tilde{\chi}$.

Suitable assumptions on $\tilde{L}$ are as follows:

- $\tilde{L} \in L^\infty([0,2\pi])$, $\|\tilde{L}\| < 2\|L_T\|_\infty$
- $\tilde{L} \geq 0$
- $\int_0^{2\pi} \tilde{L}(x)dx > \frac{1}{2}T$

The set of functions satisfying these conditions will be denoted by Λ and called admissible local times.

LEMMA 2. *For admissible perturbed local times $\tilde{L} \in \Lambda$, the perturbed bi linear form*

$$\tilde{a}(u,v) = \int_0^{2\pi} u\left(\Delta^2 + I + \tilde{L}\right) v dx \tag{33}$$

is still coercive, symmetric and continuous on $H^2_{\text{per}}([0,2\pi])$. The coercivity and continuity constants in Lemma 1 can be chosen such that α and C are the same as those for the bi linear form (29) uniformly for all $L \in \Lambda$.

The proof of Lemma 1 was performed with robustness of constants to allow for this result.

The rest of the calculation then follows standard procedure:

$$\alpha\|\tilde{u} - u\|^2_{H^2} \leq \tilde{a}(\tilde{u} - u, \tilde{u} - u)$$

Now using the PDE for the perturbations, (32), to replace the right hand side, we obtain:

$$\begin{aligned}
\alpha\|\tilde{u} - u\|^2_{H^2} &\leq \int_0^{2\pi} \frac{1}{2}(\tilde{u} - u)(\tilde{L}' - L_T') - (\tilde{u} - u)(\tilde{L} - L_T)\tilde{u}dx \\
&= \int_0^{2\pi} -\frac{1}{2}(\tilde{u} - u)'(\tilde{L} - L_T) - (\tilde{u} - u)(\tilde{L} - L_T)\tilde{u}dx \\
&\leq \frac{1}{2}\|\tilde{u} - u\|_{H^1} \cdot \|\tilde{L} - L_T\|_{L^2}\,(1 + \|\tilde{u}\|_\infty) \\
&\leq \frac{1}{2}\|\tilde{u} - u\|_{H^2} \cdot \|\tilde{L} - L_T\|_{L^2}\,(1 + \|\tilde{u}\|_\infty)
\end{aligned}$$

Dividing both sides by $\|\tilde{u} - u\|^2_{H^2}$ we obtain that this norm is either zero or at least admits the following bound:

$$\|\tilde{u} - u\|^2_{H^2} \leq C\|\tilde{L} - L_T\|_{L^2}(1 + \|\tilde{u}\|_\infty) \tag{34}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

for some $C > 0$. In just the same way the PDE (32) yields the bound

$$\alpha\|\tilde{u}\|_{H^2} \leq \|\frac{1}{2}\tilde{L} + \tilde{\chi}(\cdot; X_0, X_T) + W\|_{L^2} \tag{35}$$

Due to a Sobolev embedding we have that $\|\tilde{u}\|_\infty \leq C\|\tilde{u}\|_{H^2}$ for some suitable constant $C > 0$ so that we can use (35) to bound the term in brackets in (34). Additionally, use the fact that since $\tilde{L} \in \Lambda$ we have $\|\tilde{L}\|_\infty \leq 2\|L_T\|_\infty$ to obtain the bound in the following theorem:

THEOREM 3. *There exists a constant $C(W, \|L\|_\infty) > 0$ such that for all admissible perturbed local times $\tilde{L} \in \Lambda$ the deviation of the perturbed posterior mean $\tilde{u}$ from the unperturbed posterior mean u is bounded in the $H2$-norm:*

$$\exists C > 0 \forall \tilde{L} \in \Lambda : \ \|\tilde{u} - u\|_{H^2} \leq C(W, \|L\|_\infty)\|\tilde{L} - L\|_{L^2} \tag{36}$$

6.5. *Summary of the Analysis.* To summarize, it has been shown that the PDE for the posterior mean, (23) has a unique solution. Also, an H^2-bound for perturbations of this solution wrt. changes of the local time has been found. This enables stable numerical treatment which will be detailed in the sequel.

7. Existence and absolute continuity of the posterior. We employ the Hajek-Feldman theorem (in the version given in [13], Theorem 2.23) to establish absolute continuity. To simplify the presentation, we use the prior measure μ_0 is as given in (20). The covariance operator is

$$Q_1 = (\Delta^2 + I)^{-1}$$

and the posterior measure is Gaussian with covariance

$$Q_2 = (\Delta^2 + I + L_T)^{-1}$$

and mean $u = (\Delta^2 + I + L_T)^{-1} (\frac{1}{2}L_T' + W + \tilde{\chi}(\cdot; X_0, X_T))$ where Section 6 showed that $u \in H^2_{\text{per}}([0, 2\pi])$.

THEOREM 4. *The measures $\mu_0 = \mathcal{N}(0, Q_1)$ and $\mu = \mathcal{N}(u, Q_2)$ on $L^2(0, 2\pi)$ are absolutely continuous with respect to each other.*

CONJECTURE 1. *The operator $Q_1^{\frac{1}{2}} Q_2^{-\frac{1}{2}} - I$ is Hilbert-Schmidt on H^2_{per}.*

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

RESULT 1. *Their Radon-Nikodym derivative is given by*

(37)
$$\frac{d\mu}{d\mu_0} \propto \exp\left(\frac{1}{2}\int_0^{2\pi} \left[(L_T(a)+1)b2(a) - (L_T'(a)+\tilde{\chi}(a;X_0,X_T)+W)b(a)\right] da\right)$$

The theorem 4 will be proved in three steps as an application of general theorems on Gaussian measures on Hilbert spaces. A proof of result 1 assuming that the conjecture 1 is true will be given after that.

We proceed in three steps showing firstly that the images of the square roots of the covariance operators, $H_0 = ImQ_1^{\frac{1}{2}}$ and $ImQ_2^{\frac{1}{2}}$ agree. Secondly, we note that the difference of the respective means is an element of this image. Thirdly, we show that $(Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})(Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})^* - \mathrm{Id}$ is Hilbert-Schmidt on $L2$.

Identity of Images. Note first that we can get to the image of $Q_1^{\frac{1}{2}}$ by characterizing it as follows:

$$\begin{aligned} ImQ_1^{\frac{1}{2}} &= \left\{Q_1^{\frac{1}{2}} f | f \in L^2\right\} \\ &= \left\{g | Q_1^{-\frac{1}{2}} g \in L^2\right\} \\ &= \left\{g | \left\|Q_1^{-\frac{1}{2}} g\right\|_{L^2} < \infty\right\} \end{aligned}$$

But this norm is defined via an inner product, so that we get access to Q_1^{-1} by taking adjoints and noting that due to periodicity, boundary terms can be neglected. It is now possible to prove the claim as follows:

$$\begin{aligned} &\left\|Q_1^{-\frac{1}{2}} g\right\|_{L^2} < \infty \\ \Leftrightarrow &\int_0^{2\pi} g(\Delta^2 + I) g dx < \infty \end{aligned}$$

We now use a Hölder estimate exploiting continuity of L_T on the compact interval $[0, 2\pi]$ so that we continue the string of equivalent statements as follows:

$$\begin{aligned} \Leftrightarrow \int_0^{2\pi} g\left(\Delta^2 + I + L\right) g dx &< \infty \\ \Leftrightarrow (g, Q_2^{-1} g)_{L^2} &< \infty \\ \Leftrightarrow \left\|Q_2^{-\frac{1}{2}} g\right\|_{L^2} &< \infty \end{aligned}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

It is also clear now that the image $ImQ_1^{\frac{1}{2}}$ contains the space $H^2_{\text{per}}([0,2\pi])$.

Means are in the Range. We showed in Theorem 2 that the solution to the PDE for the posterior mean,u, is in $H^2_{\text{per}}([0,2\pi])$. We found in the preceding step of the proof that this space is contained in the image of $Q_1^{\frac{1}{2}}$ so that the difference of the means (which is just u) is as required.

Projector is only Hilbert-Schmidt different. We endeavour to show that the operator

$$(Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})(Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})^* - \text{Id} \tag{38}$$

is Hilbert-Schmidt on $L2$. By definition, this is equivalent to showing that

$$S = \sum_{n=1}^{\infty} (f_n, (Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})(Q_1^{-\frac{1}{2}}Q_2^{\frac{1}{2}})^* f_n)_{L2} - (f_n, \text{Id} f_n)_{L2} < \infty \tag{39}$$

holds for some Hilbert basis $\{f_n\}_{n=1}^{\infty}$ of $L^2(0,2\pi)$.

We consider the basis $\{e_n\}_{n=\infty}^{\infty}$ given in (22). Note that these functions are smooth. Setting $y_n = Q_1^{-\frac{1}{2}} e_n$, using the eigenvalue definition of square root of an operator it is straightforward to calculate $y_n = \sqrt{1+n^4}e_n$. We can then reformulate (39) as follows:

$$\begin{aligned} S &= \sum_{n=-\infty}^{\infty} (Q_1^{-\frac{1}{2}} e_n, Q_2 Q_1^{-\frac{1}{2}} e_n)_{L^2} - (e_n, e_n)_{L^2} \\ &= \sum_n (y_n, Q_2 y_n)_{L^2} - (Q_1^{\frac{1}{2}} y_n, Q_1^{\frac{1}{2}} y_n)_{L^2} \\ &= \sum_n (y_n, (Q_1 - Q_2) y_n)_{L^2} \end{aligned}$$

Since $Q_2 = \left(\Delta^2 + I + L_T\right)^{-1}$ is the left hand side operator in the PDE for the posterior mean, (30), we know its inverse - the differential operator $\left(\Delta^2 + I + L_T\right)$; note that the domain is unimportant since we are dealing with smooth functions e_n and y_n throughout.

$$\begin{aligned} Q_2^{-1} y_n &= \left(\Delta^2 + I + L_T\right) y_n = (n^4+1) y_n + L_T y_n \\ \Rightarrow y_n &= (n^4+1) Q_2 y_n + Q_2 (L_T y_n) \\ \Rightarrow Q_2 y_n &= Q_1 y_n - \frac{1}{n^4+1} Q_2 (L y_n), \end{aligned} \tag{40}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

where we have exploited the fact that the basis (22) is an eigenbasis of Q_1 with eigenvalues $Q_1 y_n = \frac{1}{n4+1} y_n$ in the last line.

So that the sum S can be continued as follows:

$$\begin{aligned} S &= \sum_n n^4 \langle y_n, \frac{1}{n^4+1} Q_2(L_T y_n)\rangle \\ &= \sum_n \frac{1}{n^4+1} \langle Q_2 y_n, L_T y_n\rangle \end{aligned}$$

Here, symmetry of Q_2 has been exploited. Next, we use the relation (40) again to substitute for $Q_2 y_n$ and we obtain:

$$\begin{aligned} S &\leq \sum_n \left(\frac{1}{n^4+1} |(Q_1 y_n, L_T y_n)_{L^2}| + \frac{1}{n^4+1} |(Q_2(L_T y_n), L_T y_n)_{L^2}| \right) \\ &\leq \sum_n \frac{1}{n^4+1} \left(\frac{1}{1+n^4} \|y_n\|_{L^2}^2 \|L_T\|_\infty + \|L_T\|_\infty^2 \|Q_2\| \frac{\|y_n\|_{L^2}^2}{n^4+1} \right), \end{aligned}$$

where we have used that Q_2 is a bounded operator and again we have used Hölder's inequality to eliminate L_T. Now note that $\|y_n\|_{L^2}^2 = 1 + n^4$ which is straightforward to verify remembering that e_n are eigenfunctions of Q_1. Overall we obtain

$$\begin{aligned} S &\leq \sum_{n=-\infty}^{\infty} \frac{1}{n^4+1} \left(\|L_T\|_\infty + \|L_T\|_\infty^2 \|Q_2\| \right) \\ &< \infty \end{aligned}$$

□

It turns out that the sum S converges as long as the perturbation that turns Q_1 into Q_2 is an inverse differential operator of lower order than Q_1 – if its order is only lower by one, summability of the sequence is not obtained and one has to exploit that only finite Hilbert-Schmidt norm is required (rather than the trace class property we showed here).

To summarize, all three hypotheses of the Hajek-Feldman theorem from [13] have been shown to hold so that absolute continuity of the posterior measure with respect to the prior measure has been shown.

7.1. *Identifying the Radon-Nikodym derivative.* In this subsection we continue work on the prior and posterior measures and prove Result 1 subject to Conjecture 1, i.e. we demonstrate that the measures' Radon-Nikodym derivative is indeed identical with the likelihood (4) as intended. To this end,

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

we proceed in two steps, first multiplying a random variable with distribution given by the prior measure by the correct linear operator to obtain a centred Gaussian random variable with the desired covariance operator. In a second step we translate this random variable to obtain the desired mean. Throughout these two steps we keep track of the prefactors introduced and finally identify the product of those prefactors as the desired likelihood up to constants of integration.

7.1.1. *Covariance: Multiplication by $I + K$.* We multiply the random variable

$$X \sim \mu_0$$

which is distributed according to the prior measure by the operator $T = Q_1^{\frac{1}{2}} Q_2^{-\frac{1}{2}}$. We intend to use theorem 6.4.5. from [6] which requires that the operator $K = T - I$ be Hilbert-Schmidt on $H(\mu_0) = H^2_{\text{per}}$. This requirement is more stringent than the fact established above that $CC^* - I$ is Hilbert-Schmidt. We do not show this fact, but state it as conjecture 1.

Now for $S = T^{-1}$, theorem 6.4.5. from [6] guarantees that

$$\frac{d\mu_0 \circ S^{-1}}{d\mu_0} \propto \exp\left(\delta K(b) - \frac{1}{2}|K(b)|^2_{H2_{\text{per}}}\right),$$

where $\delta K(b)$ denotes the divergence of the vector fields Kb with respect to the Gaussian measure μ_0 and we have left out constants of proportionality that do not depend on b. To calculate this factor, note that by virtue of theorem 5.8.4 in [6] we may as well assume the divergence to be taken of the symmetrized operator $K_s = \frac{1}{2}(K + K^*)$. We will first perform the calculation assuming that b is smooth. Then the following manipulations

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

hold:

$$\begin{aligned}
&\delta K_s(b) - \frac{1}{2}|Kb|^2_{\mathcal{H}} = \mathrm{trace}(K_s) - (b, K_s b) - \frac{1}{2}(Kb, Kb)\\
=&C + \frac{1}{2}(b,b) - \frac{1}{2}(b,b) - (b, K_s b) - \frac{1}{2}(Kb, Kb)\\
=&C + \frac{1}{2}(b,b) - \frac{1}{2}\left((I+K)b, (I+K)b\right)\\
=&C + \frac{1}{2}\int_0^{2\pi} (\Delta b(a))^2 + b^2(a) da\\
&- \frac{1}{2}\int_0^{2\pi} \left(\left[\sqrt{(\Delta^2+I)^{-1}}\sqrt{\Delta^2+I+L_T}b\right]''\right)^2 + \left[\sqrt{(\Delta^2+I)^{-1}}\sqrt{\Delta^2+I+L_T}b\right]^2 da\\
=&C + \frac{1}{2}\int_0^{2\pi} (\Delta b(a))^2 + b^2(a) da\\
&- \frac{1}{2}\int_0^{2\pi} \left(\sqrt{\Delta^2+I+L_T}b\right)\cdot\left(\sqrt{(\Delta^2+I)^{-1}}\left(\frac{\partial^4}{\partial a^4}+I\right)\sqrt{(\Delta^2+I)^{-1}}\right)\sqrt{\Delta^2+L_T}b da\\
=&C + \frac{1}{2}\int_0^{2\pi} (\Delta b(a))^2 + b^2(a) da - \frac{1}{2}\int_0^{2\pi}\left((\Delta b)^2 b^2(a) + L_T b^2\right) da\\
=&C - \frac{1}{2}\int_0^{2\pi} L_T(a) b(a)^2 da,
\end{aligned}$$

where any unlabelled inner products refer to $\mathcal{H}$ and we have used a constant $C \in \mathbb{R}$. Considering only the start and end of these calculations we can now extend them to all of L^2 by continuity considering the measurable extensions $\widehat{K}$ where needed. We thus find that the random variable TX has a Gaussian distribution and is absolutely continuous with respect to X with Radon Nikodym derivative

$$\frac{d\mathcal{L}(TX)}{d\mathcal{L}(X)} \propto \exp\left(-\frac{1}{2}\int_0^{2\pi} L_T(a) b(a)^2 da\right). \tag{41}$$

Since T is a closed operator it is easy to see that the expected value of TX is still zero. The Covariance is obtained as follows:

$$\begin{aligned}
\mathbb{E}TX(TX)^* &= \mathbb{E}\sqrt{\Delta^2+I+L_T}^{-1}\sqrt{\Delta^2+I}XX*\sqrt{\Delta^2+I}\sqrt{\Delta^2+I+L_T}^{-1}\\
&= \sqrt{\Delta^2+I+L_T}^{-1}\sqrt{\Delta^2+I}(\Delta^2+I)^{-1}\sqrt{\Delta^2+I}\sqrt{\Delta^2+I+L_T}^{-1}\\
&= (\Delta^2+I+L_T)^{-1}
\end{aligned}$$

which is as desired.

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

7.1.2. *Mean: Translation.* We translate the measure $\mathcal{L}(TX)$ by considering the random variable

$$Y = TX + h$$

where h is given as the solution of the PDE (30),

$$(\Delta^2 + I + L_T)h = \frac{1}{2}L'_T + \tilde{\chi}(\cdot; X_0, X_T) + W$$

Note that since the operator on the left hand side of this PDE is exactly the inverse of the covariance operator of the random variable TX, this h fulfils the requirement of lying inside the Cameron-Martin space of the measure $\mathcal{L}(TX)$ so that Corollary 2.4.3 from [6] yields absolute continuity of the translated measure $\mathcal{L}(Y)$ and gives the Radon-Nikodym derivative as

$$\frac{d\mathcal{L}(Y)}{d\mathcal{L}(TX)} = \exp\left(\int_0^{2\pi}\left[(\Delta^2 + I + L_T)h\right](a)b(a)da - \frac{1}{2}\|h\|^2_{\mathcal{H}}\right) \tag{42}$$

We can neglect the second summand in the exponent since it does not depend on b (we can think of it as a normalizing constant). It is clear that TX and Y have the same covariance operators and that the expected value has been shifted so that $\mathbb{E}Y = h$.

7.1.3. *Combining the Radon-Nikodym factors.* Going from the random variable X to the random variable Y we have accumulated two Radon-Nikodym factors given by (41) and (42) respectively. Multiplying these two factors together we see that since $\mathcal{L}(X) = \mu_0$

$$\frac{d\mathcal{L}(Y)}{d\mu_0} \propto \exp\left(\frac{1}{2}\int_0^{2\pi}\left(L_T(a)b(a)^2 - \left[L' + \tilde{\chi}(\cdot; X_0, X_T) + W\right](a)b(a)\right)da\right) \tag{43}$$

as desired for (4) to hold.

7.2. *Rigorous Bayesian Framework.* Writing down conditional probability measures like

$$\mathbb{P}(b|\{x_t\}_{t=0}^T)$$

can be problematic as we condition on an event of measure zero. The existence of these conditional expectations turns out to be rather subtle; in the given context we can, however, write down regular conditional probability measures explicitly. To introduce a rigorous Bayesian framework, one can

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

first define a product measure on the product space of path observations and drift functions and then multiply this measure by the desired likelihood to establish existence of a probability measure on the product space. It is then a straightforward matter to verify that the required marginal and conditional densities of that measure are as desired. For the sake of brevity, we shall not give details.

8. Estimating L_T from discrete high-frequency observations.

8.1. *General Remarks.* In practice, the exact function L_T is not available, whereas what typically is available are observations of the diffusion process $\{X_{t_i}\}_{i=1}^N$. We will consider the case of high-frequency observations, i.e. we assume constant final time T, and that $N \to \infty$ and $\max_{i\in 1,\dots,N_1}(t_{i+1}-t_i) \to 0$. From the paper of Jacod [16] we can then get *pointwise* estimates of L_T, i.e. he constructs random variables $U(h)_T^n$ such that

$$\mathbb{P}\left(|U_T^n(h) - \lambda(H_h)L_T(0)| > \varepsilon\right) \to 0. \tag{44}$$

Here, h is some cutoff function, in the case of a histogram-like counting approach this could be a step function centred at 0 with some fixed width δ (corresponding to the bin width at that point). This estimator is formulated for the point value of L_T at 0 and there are uniformity statements as T is varied. Additionally, there are statements about the distribution obtained when the above difference between U and L_T is scaled by $\sqrt{n}$, somewhat akin to asymptotic normality. However, there is **no** statement about how estimates of L_T at different points all obtained from the same time series behave, so obtaining statements for the L^2-norm of the error is not straightforward.

8.2. *post-factum local times.* One way to approach this is to combine Hölder-Continuity of the local time with pointwise L^2-estimates as follows:

Let $\hat{L}_T^{x_i} = L_T^{x_i} + e_i$ be an estimate of the local time at the point $x_i \in [0, 2\pi]$ with error e_i. We extend this estimate simply to a constant function $\hat{L}_T^{\cdot}$ on the interval $[x_i, x_{i+1})$ and an estimate on the whole of $[0, 2\pi)$ is obtained by piecing these intervals together.

Using results from Jacod's paper [16] we know that each e_i converges to zero in probability. Now consider the $L2$ norm of the error on one sampling

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

interval:

$$\begin{aligned}\int_{x_i}^{x_{i+1}} |L_T^{x_i} + e_i - L_T^x|^2\, dx &\le \int_{x_i}^{x_{i+1}} e_i^2 + 2|e_i||L_T^{x_i} - L_T^x| + |L_T^{x_i} - L_T^x|^2 dx \\ &\le 2e_i^2|x_{i+1} - x_i| + 2\int_{x_i}^{x_{i+1}} |L_T^{x_i} - L_T^x|^2 dx \\ &\le 2e_i^2|x_{i+1} - x_i| + 2C\int_{x_i}^{x_{i+1}} |x_i - x|^{2\alpha} dx \\ &= 2e_i^2|x_{i+1} - x_i| + \frac{2C}{2\alpha+1}|x_{i+1} - x_i|^{2\alpha+1}\end{aligned}$$

where we have used that the local time $L_T^{\cdot}$ is Hölder with exponent α and Hölder-constant C. To simplify, assume that an equispaced grid is used, i.e. $x_{i+1} - x_i = \frac{2\pi}{M}$ for some $M \in \mathbb{N}$ which may be large. Summing all the error contributions over the intervals $i = 1, \dots, M$ we then obtain the following bound for the L^2-error:

$$\begin{aligned}\|\hat{L}_T^{\cdot} - L_T^{\cdot}\|_{L^2(0,2\pi)} &\le \sum_{i=1}^{M} 2e_i^2|x_{i+1} - x_i| + \frac{2C}{2\alpha+1}|x_{i+1} - x_i|^{2\alpha+1} \\ &\le 4\pi \max_{i=1,\dots,M} e_i 2 + \frac{2C}{2\alpha+1}\sum_{i=1}^{M}\left(\frac{2\pi}{M}\right)^{2\alpha+1} \\ &= 4\pi \max_i e_i^2 + \frac{2^{2\alpha+2}\pi^{2\alpha+1}}{2\alpha+1} C M^{-2\alpha}.\end{aligned}$$

The trouble with this estimate is that M must be chosen large enough to compensate for C - but C is not known and depends on the continuous time sample path. Unless a-priori estimates of the Hölder constant of continuity are available, this method can only be made to work if M is chosen large enough as a function of the samplepath. Thus, overall convergence in probability can only be obtained if C can be bounded using only the high frequency data.

8.3. *Admissible Perturbations.* The restriction on admissible perturbed local times $\tilde{L} \in \Lambda$ is such that, in particular, local time functions L_{T_2} taken at a later time $T_2 > T$ where $T_2 - T$ is sufficiently small are admissible as perturbed local times $\tilde{L}$. This is due to the *joint* continuity of local times, $L_T(x)$, in x and T jointly. The admissible local times will not include all later times since the L_∞-norm of L_T will grow too big eventually. For any particular finite final time, this can be fixed however by choosing α small enough to do the trick for both local times in question which is always possible.

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

9. Discussion. This paper is a first step towards Bayesian non-parametric drift estimation for Itô SDEs, providing a rigorously analyzed simple case as well as various numerical extensions of the proposed methodology.

Many of the limitations imposed are of a technical nature, so that several extensions seem likely possible:

- Treat Itô SDEs on the whole real line by considering appropriate function spaces weighted with an invariant measure subject to ergodicity assumptions.
- Extend the methodology to higher dimensions. Dimensions two and three would be expected not to yield major complications, higher dimensions make discretizing the PDE uncomfortable and very sparse grids would have to be used. Alternatively, methods to select base functions to arrive at a reduced spectral representation as used in quantum chemistry seem interesting.
- The use of the prior inverse covariance operator $\mathcal{A} = \eta\Delta^2 + \epsilon I$ is largely illustrative, and it would be natural to consider other prior structures (Note that Δ as used in Section 1.4 can be modified to impose a Markov property on prior and posterior which could be natural in some applications.) From a rigorous point of view, it should be noted that each such generalization leads to extra technical conditions to be checked.

Further work will consider genuinely discrete time observations (i.e. renounce the high frequency assumption). We will investigate the use of an MCMC data-augmentation scheme involving sampling alternatingly from local times given discrete time observations and drift functions.

In the discrete-time case, an interesting extension considers unknown (and possibly in-homogeneous) volatility for the observation process. There are natural methodologies for dealing with this problem which are developed in the Bayesian parametric literature (see for example [20]).

REFERENCES

[1] G.O. Roberts A. Beskos, O. Papaspiliopoulos. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

[2] R. A. Adams. *Sobolev Spaces*. Academic Press, 1975.

[3] Federico M. Bandi and Peter C. B. Phillips. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283, 2003.

[4] Alexandros Beskos, Omiros Papaspiliopoulos, Gareth O. Roberts, and Paul Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(3):333–382, 2006. With discussions and a reply by the authors.

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

[5] J. P. N. Bishwal. *Parameter estimation in stochastic differential equations*, volume 1923 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008.

[6] V. I. Bogachev. *Gaussian Measures*. AMS, 1998.

[7] D. Braess. *Finite Elemente, Schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer, 1997.

[8] H. Brézis. *Analyse fonctionelle – Théorie et applications*. Dunod, 1999.

[9] Fabienne Comte, Valentine Genon-Catalot, and Yves Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007.

[10] B.R.Brooks et al. Charmm: A program for macromolecular energy, minmization and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[11] L. C. Evans. *Partial Differential Equations*. AMS, 1998.

[12] A. W. van der Vaart F. H. van der Meulen and J. H. van Zanten. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, 12(5):863–888, 2006.

[13] J. Zabczyk G. Da Prato. *Stochastic Equations in Infinite Dimensions*. CUP, 1992.

[14] A. van der Vaart H. v. Zanten. Rates of contraction of posterior distributions based on gaussian process priors. *Annals of Statistics*, 36:1435–1463, 2008.

[15] W. Hackbusch. *Elliptic differential equations : theory and numerical treatment*. Springer, 1992.

[16] J. Jacod. Rates of convergence to the local time of a diffusion. *Annales de L'institut Henri Poincaré, Probabilités et statistique*, 34:505–544, 1998.

[17] S. Chib O. Elerian and N. Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993, 2001.

[18] N. G. Polson and G. O. Roberts. Bayes factors for discrete observations from diffusion processes. *Biometrika*, 81(1):11–26, 1994.

[19] J.N. Reddy. *An Introduction to the Finite Element Method*. McGraw-Hill, 1984.

[20] G. O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[21] J. C. Robinson. *Infinite-dimensional Dynamical Systems*. CUP, 2001.

[22] T. Schlick. *Molecular Modeling and Simulation, an Interdisciplinary Guide*. Springer, New York, 2002.

[23] Y.A.Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004.

10. Appendix A: Technical Proofs.

10.1. *Existence of the Prior Measure.* To show that the prior is well-defined in (20), we note that the operator $\mathcal{A} = \Delta^2 + I$ with domain $\mathcal{D}(\mathcal{A}) = H^4_{per}$ is positive definite and symmetric.

Its inverse is compact since H^4_{per} is compactly embedded in $L^2(0, 2\pi)$ by the Rellich-Kondrachov Compactness Theorem (see e.g. [11], Chapter 5.7). Thus, the operator $\mathcal{A}^{-1}$ admits an orthonormal basis of eigenfunctions $\{e_k\}_{k\in\mathbb{Z}}$ (a standard result that can be found e.g. in [8], chapter VI.4) which

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

can, in this case, be given explicitly as in (22). The associated eigenvalues are $\lambda_k = \frac{1}{k^4+1}$. Furthermore, the operator $\mathcal{A}^{-1}$ is trace class since

$$\sum_{k=-\infty}^{\infty} \int_0^{2\pi} e_k(a)(\mathcal{A}^{-1}e_k)(a)da = \sum_{k=-\infty}^{\infty} \frac{1}{k^4+1} < \infty.$$

Since the desired covariance operator $\mathcal{A}^{-1}$ is thus symmetric, non-negative and trace-class, Theorem 2.3.1 from [6] guarantees the existence of a corresponding Gaussian measure.

10.2. *Proof of Lemma 1.* In this part we prove Lemma 1.

Proof:

Linearity and symmetry of the bi linear form are clear from inspection of (25). In order to prove coercivity we proceed in stages. Firstly, the operator is split in two parts which will be dealt with in turn:

$$a(u,u) = \int_0^{2\pi} (\Delta u)^2 + u^2 G dx = a_1(u,u) + a_2(u,u)$$

where the two parts are given as

$$a_1(u,u) = \int_0^{2\pi} (\Delta u)^2 dx$$
$$a_2(u,u) = \int_0^{2\pi} u^2 G dx$$

First part of the Proof: a_1. In this part of the proof, we use the Poincaré inequality to get a lower bound on the H^2- and H^1-seminorm content of the quadratic form.

In order to be able to use the Poincaré inequality, introduce the function

$$v(x) = u'(x) - u'(0)1(x).$$

Note that $v \in H_0^1([0,2\pi])$ since $u'(0) = u'(1)$ due to periodicity (and having chosen to consider a continuous version of u'). Now consider the first integral in the above inequality:

$$\begin{aligned}\int_0^{2\pi} (\Delta u(x))^2 dx \geq \int_0^{2\pi} (\Delta u)^2 dx &= \int_0^{2\pi} (\Delta u - v' + v')^2 dx \\ &= \int_0^{2\pi} (\Delta u - v')^2 + 2(\Delta u - v')v' + (v')^2 dx\end{aligned}$$

Now note that

$$\Delta u(\cdot) - v'(\cdot) = \left(u'(\cdot) - u'(\cdot) + u'(0)1(\cdot)\right)' = \left(u'(0)1(\cdot)\right) = 0.$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

So that we continue the string of inequalities from above by applying the Poincaré inequality $|v|_{L^2} \leq C'|v'|_{L^2}$ exploiting the fact that $v \in H_0^1$:

$$\int_0^{2\pi} (\Delta u)^2 dx = \int_0^{2\pi} (v')^2 dx \geq \frac{1}{C'} \int_0^{2\pi} v^2 dx$$
$$= \frac{1}{C'} \int_0^{2\pi} \Big((u'(x))^2 - u'(0)u'(x) + (u'(0))^2 1(x) \Big) dx$$
$$= \frac{1}{C'} \int_0^{2\pi} \left((u')^2 + \underbrace{(u'(0))^2}_{\geq 0} \right) dx - \underbrace{2 \int_0^{2\pi} u'(x) dx}_{=2(u(2\pi)-u(0))=0}$$
$$\geq \frac{1}{C'} \int_0^{2\pi} (u'(x))^2 dx,$$

where we have used periodicity of u to cancel the last expression. Splitting the integral $\int_0^{2\pi} (\Delta u)^2 dx$ into two equal parts and applying the above inequality to only one of the two parts we have the following inequality:

$$a_1(u,u) \geq \frac{1}{2} \int_0^{2\pi} (\Delta u)^2 dx + \frac{1}{2C'} \int_0^{2\pi} (u')^2 dx. \tag{45}$$

While the analytical treatment is a little involved, in principle the only thing we have done is to eliminate the linear, non-constant functions from the kernel of the linear operator associated with the quadratic form a_1.

Second Part of the Proof: a_2. In this part of the proof we again use Poincaré's inequality to borrow some H^1-component of u from part one and use it to compensate some L^2-deficiency in the second part. Overall an estimate of the L^2-content of the quadratic form from below can be performed for very general classes of 'nearby' local times in one go.

In the same way setting up for exploiting the Poincaré inequality, let us define

$$\bar{u} = u(0)1(x)$$

Again, having chosen a continuous version of u we note that due to periodicity $u(0) = u(2\pi)$ and so $u - \bar{u} \in H_0^1([0, 2\pi])$.

So let us start with the second part and introduce the difference $u - \bar{u}$ as

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

follows:

$$\int_0^{2\pi} u^2(x)G(x)dx = \int (u-\bar{u}+\bar{u})^2 G dx$$
$$\int \left[(u-\bar{u})^2 + 2(u-\bar{u})\bar{u} + \bar{u}^2\right] G dx$$
$$\geq \int \left[(1-\delta)(u-\bar{u})^2 + \left(1-\frac{1}{\delta}\right)\bar{u}^2\right] G dx.$$

To obtain the last line we have used the Cauchy-Schwarz inequality with δ, namely $|2ab| \leq \delta a^2 + \frac{1}{\delta}b^2$. The only constraint on the parameter in this case is $\delta \in (1,\infty)$; we will fix δ later.

Since $(1-\delta)$ is negative, we need to 'borrow' some $\|u\|_{L^2}$ from the first part, namely (45) to compensate for this. Starting from the second term on the right hand side of that inequality, we proceed as follows:

$$\frac{1}{2C'}\int (u')^2 dx = \frac{1}{2C'}\int \left((u-\bar{u})'\right)^2 dx$$
$$= \frac{1}{4C'}\int (u')^2 dx + \frac{1}{4C'}\int ((u-\bar{u})')^2 dx,$$

where we have used that $\bar{u}$ is actually a constant! We now leave the first term for the coercivity estimate, to retain some $|u|_{H^1}$-seminorm component, and add the second term to a_2 to compensate as outlined above:

$$A := a_2(u,u) + \frac{1}{4C'}\int ((u-\bar{u})')^2 dx \geq a_2(u,u) + \frac{1}{4C'^2}\int (u-\bar{u})^2 dx \tag{46}$$
$$\geq \int (1-\delta)(u-\bar{u})^2 G + \left(1-\frac{1}{\delta}\right) G\bar{u}^2 + \frac{1}{4C'^2}(u-\bar{u})^2 dx.$$

Here, we have used the Poincaré inequality. Now, we will estimate L by its ∞-norm, which is possible since L is continuous on the compact set $[0,2\pi]$:

$$A \geq \int \left(\frac{1}{4C'^2} - (\delta-1)\|G\|_\infty\right)(u-\bar{u})^2 + \left(1-\frac{1}{\delta}\right)\bar{u}^2 L_T dx$$
$$= \left(\frac{1}{4C'^2} - (\delta-1)\|G\|_\infty\right)\|u-\bar{u}\|^2_{L^2([0,2\pi])} + (\bar{u}(0))^2\left(1-\frac{1}{\delta}\right)\int_0^{2\pi} G(x)dx.$$

Now note that G is continuous on a compact domain. Hence, its integral over $[0,2\pi]$ is finite and we denote it by T. In applications, $G = L_T + 1$ is the usual case, so that to be able to tackle perturbed local times in the sequel, we allow an extra 'safety' factor of two both for the integral of G and the $\|G\|_\infty$-norm, so that we obtain the following inequality:

$$A \geq \left(\frac{1}{4C'^2} - 2(\delta-1)\|G\|_\infty\right)\|u-\bar{u}\|^2_{L^2([0,2\pi])} + \frac{T}{4\pi}\left(1-\frac{1}{\delta}\right)\|\bar{u}\|^2_{L^2([0,2\pi])}.$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

We now wish to guarantee that the first prefactor, $\frac{1}{4C'^2} - 2(\delta-1)\|L\|_\infty$, stays positive. A brief calculation shows that this can be achieved by choosing

$$\delta \in \left(1, 1 + \frac{1}{8C'^2\|G\|_\infty}\right).$$

For such a choice of δ define the new constant γ as follows:

$$\gamma = \min\left\{\frac{1}{4C'^2} - (\delta-1)\|G\|_\infty, \frac{T}{4\pi}\left(1 - \frac{1}{\delta}\right)\right\}.$$

We then have

$$\begin{aligned} A &\geq \gamma\left(\|u-\bar{u}\|^2_{L^2([0,2\pi])} + \|\bar{u}\|^2_{L^2([0,2\pi])}\right) \\ &\geq \frac{\gamma}{2}\|u\|^2_{L^2([0,2\pi])}, \end{aligned}$$

where we have used the triangle inequality to get the L^2-norm of u.

Third Part of the Proof: Synthesis. In this part of the proof we compose the results obtained in parts one and two and define the coercivity constant α.

The first part of the proof established that

$$a_1(u,u) \geq \frac{1}{2}\int_0^{2\pi}(\Delta u)^2 dx + \frac{1}{2C'}\int_0^{2\pi}(u')^2 dx.$$

Recalling the definition of the quantity A, namely (46), part two of the proof established that

$$A := a_2(u,u) + \frac{1}{4C'}\int((u-\bar{u})')^2 dx \geq \frac{\gamma}{2}\|u\|^2_{L^2([0,2\pi])}$$

for some positive γ that does not depend on u. Furthermore, the constants are uniform for perturbed local times $\tilde{L}$ such that $\int_0^{2\pi}\tilde{L}(x)dx \geq \frac{T}{2}$ and $\|\tilde{L}\|_\infty \leq 2\|L_T\|_\infty$. Using these two estimates we obtain:

$$\begin{aligned} a(u,u) &= a_1(u,u) + a_2(u,u) \\ &\geq \frac{1}{2}\int_0^{2\pi}(\Delta u)^2 dx + \frac{1}{2C'}\int_0^{2\pi}(u')^2 dx + a_2(u,u) \\ &= \frac{1}{2}\int_0^{2\pi}(\Delta u)^2 dx + \frac{1}{4C'}\int_0^{2\pi}(u')^2 dx + A \\ &\geq \frac{1}{2}\int_0^{2\pi}(\Delta u)^2 dx + \frac{1}{4C'}\int_0^{2\pi}(u')^2 dx + \frac{\gamma}{2}\|u\|^2_{L^2([0,2\pi])}. \end{aligned}$$

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism

Now set the coercivity constant as

$$\alpha = \min\left\{\frac{1}{2}, \frac{1}{4C'}, \frac{\gamma}{2}\right\}$$

to finally obtain the estimate from below:

$$\begin{aligned} a(u,u) &\geq \alpha\left(\int_0^{2\pi}(\Delta u)^2 dx + \int_0^{2\pi}(u')^2 dx + \int_0^{2\pi} u^2 dx\right) \\ &= \alpha\|u\|^2_{H^2([0,2\pi])}. \end{aligned}$$

This completes the proof of coercivity.

Showing continuity is comparatively easy:

$$\begin{aligned} a(u,v) &= \int \Delta u \Delta v + uGv dx \\ &\leq |u|_{H^2}|v|_{H^2} + \|G\|_\infty |u|_{L^2}|v|_{L^2} \leq \max\left\{1, 2\|G\|_\infty\right\}\|u\|_{H^2}\|v\|_{H^2}. \end{aligned}$$

Again, we have left a 'safety' factor of two to make a larger class of perturbed local times $\tilde{L}$ admissible in applications of this lemma. □

CRiSM Paper No. 09-29, www.warwick.ac.uk/go/crism